ENACTING MEMORY
Rethinking our Capacity to Remember

Richard Fitzpatrick

Master of Science
Cognitive Science
School of Informatics
The University of Edinburgh
2016

Abstract

Current work into how our declarative memory system operates has begun to overturn our intuitive notions of what our memory is for. There is enough evidence to suggest our memory for events and facts arose from our need to navigate through our surroundings efficiently, and utilises the very same underlying neural architecture. Enactivist accounts of cognition strive to focus our attention onto the interaction of an organism with its environment, aiming to reduce the need to refer to internal representational vehicles, as well as highlighting the importance of the organism itself in cognitive processes. Autopoietic Enactivism allows for the development of a description of declarative memory which fits with empirical data and supplies a unique insight into both memory phenomena and the underlying processes. This paper provides such a sketch of this enactive memory system, emphasising the continuity between navigation and episodic memory, the key functions of gist extraction and hypothesis testing, and the potential extension of this description into semantic memory understanding.

Acknowledgements

I would like to thank firstly my supervisor Dr Dave Ward for allowing me to indulge in this long held pet project of mine, and for his support and guidance throughout. I would also like to thank my friends and family for putting up with all my discussions of this work, especially those in 1 George Square who had to endure me trying to overturn their philosophies about their own work. Last but not least, thank you Jamie for putting up with me and for keeping me sane through this whole process.

Declaration

I hereby declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified. I have read and fully understood the University's plagiarism guidelines.

Richard Fitzpatrick

Contents

0. Introduction

A man's memory is not a summation; it is a chaos of vague possibilities.
Jorge Luis Borges

Enactivism and memory on the surface don't mix. How can a view which eschews traditional concepts of representation and objectivity be capable of explaining a cognitive process which by its very terminology suggests these things? This paper is an attempt to show that an enactivist stance, whilst difficult to bring together, is capable of explaining declarative memory in a manner which at times surpasses the explanatory capabilities of more cognitivist or connectionist flavours. From the onset, I do not claim that enactivism is in some way ready or able to supplant existing theoretical paradigms. Nor do I intend my claims to be regarded as final products; this is an area of research ripe for study, but with few people doing so.

Instead, I intend to sketch out a description of enactivism, memory, and (to a lesser extent) representation, which, when taken together, provide a different vista on the wealth of memory research within the sciences and beyond. Each may be accepted as individuals or as a whole; what is important is that a multidisciplinary assessment of the field highlights that a cognitivist take on memory is old-fashioned and no longer the default position. We just, as of yet, haven't quite got round to realising it.

To flesh out this position I will be splitting the paper into four parts. Just what I mean by 'enactivism', 'memory', and 'representation' are the focus of **Section I**. From there, **Section II** builds on the idea that memory does not function primarily as a process for remembering. By appeal to work within neuroscience on spatial memory in rodents, we shall see that the roots of memory are wholly active. In addition, by appeal to memory work in general, the notion of representation as is traditionally conceived will be questioned. **Section III** constitutes an attempt to cohere this work into a working account of memory which is rooted in enactive concepts. I will then mount a challenge from a cognitivist and a connectionist position, and assess the relative strengths of each before concluding in **Sections IV** and **V**.

I. Defining Our Terms

I.I A Short History of Cognitivism, Connectionism and Enactivism

Cognitive science might be thought of being in its Third Age. The First, heralded by work such as Simon and Newell's in 1976, ushered in the traditional position of cognitivism. The Second arrived in 1985, with the publication of **PDP** by Rumelhart and McClelland. The Third's beginning is debatable. Suffice to say it is the current Age, and one in which traditional, connectionist, dynamical, enactivist and a whole host of other '-ists' and '-isms' vie for supremacy over the rest.

Each of these distinct positions arose from the pioneering work of cyberneticists such as Ashby in the mid-20th century (Varela, Thompson & Rosch, 1991). That each in turn has been adopted rather than together is likely due to desire for an easy single unifying framework for cognition. Nevertheless, it is becoming clear that this desire is a fool's errand. What follows is an attempt to frame our current question by embedding it within the history of this fertile and febrile field.

I.I.i Cognitivism

Symbols are at the core of cognitivism. Thought is the manipulation of symbols; the *syntax,* not the semantics, are what's important. Much like our standard notion of a computer, a cognitive system in this sense would take an input, and process it using manipulation of these symbols. Once done, the system produces an output, which can be expressed in the same symbols. Arising as it did around the decline of behaviourism, there is a desire to explain as much as possible of cognition through these symbolic manipulation steps. The most well-known current holder of this position is Jerry Fodor, who has championed the idea of a Language of Thought (see e.g. Fodor, 1975). Such a language consists of types and tokens, which are subsequently combined or divided in ways that can elicit our cognitive behaviour.

Cognitivism in this traditional sense is little supported nowadays, both as a philosophical position and within closely related fields such as neuroscience. There is still however a great deal of the surrounding assumptions of the mind generated by this field still in place. We may not talk of symbolic manipulations at the sub-personal level of cognition, but it is commonplace to still utilize the idea of step-wise processing and representations which carry content.

I.I.ii Connectionism

The difference between cognitivism and connectionism can be summed up in one word-*dynamics*. A connectionist may maintain the status of representations, and of the brain as a control centre. Yet at no point within a connectionist model is there the requirement for any symbolic processing. What exists instead is an ongoing, constantly updating set of differential equations which determine the state of a system or systems at any given time *t.* These states are distributed across different brain regions and neuronal populations (or in experimental work, 'nodes'), and as such do not always fit into a functionalist architecture. This achieves a certain flexibility to the processes which are possible in any sufficiently complex connectionist system.

At no point are any of the nodes of this network *representing* the features in a discrete manner. The features are emergent, but are only really brought into being by our own observation and interpretation of the computational goings-on. This is why in some cases connectionism is seen as a sub-symbolic paradigm; if there are representations, they are at any given time a mixture of a number of cognitivist 'symbols'. Put more clearly, connectionism states processing of information is massively parallel. There is no processing of one representation, then the next, and the next etc. A node is processing and integrating at any one time a mass of information which it then sends on to connecting nodes.

Thus any representations we glean from this process can be argued to be symbolic only by virtue of how we interpret it. There is little need to really worry about the status of representations as a connectionist outwith philosophy. The concerns that might arise are similar to ones we outline later, but are not central tenets of the connectionist dogma.

I.I.i Enactivism

Enactivism makes its claims upon providing an explanatory paradigm for cognition by avoiding representational talk entirely. In its earliest form, enactivism situated itself as the 'middle way' between objective external analyses of cognition and considered phenomenological reflection (Varela, Thompson & Rosch, 1991). The focus was to see any cognitive agent as being no different to the environment to which it was currently occupying. The agent generates the environment from the physical features and its own internal dynamical processing. The coupling of the two, the *interaction*, becomes the domain of thought. The mind is unseated from its usual independent overseer role, processing the external world and generating outputs. Now it exists only as long as there is activity going on- in the brain and elsewhere.

Perception is no longer passive, awaiting input, it actively seeks it. That is to say, perception is merely a form of action, one that improves the accuracy of future responses.

The early understanding of perception/action links as some sequential processing with input then output begins to sound difficult to reconcile when taken into situations where thinking would have to occur so fast as to bypass any waiting period of processing:

> Imagine that you, a tennis player, are standing just behind your deuce corner's baseline. A ball is served to your forehand — you pivot (or rotate) so that your side is to the ball's incoming path and start to take your racket back for the forehand return. Keep visualizing up to where you're about halfway into the stroke's forward motion; the incoming ball is now just off your front hip, maybe six inches from point of impact. Consider some of the variables involved here. On the vertical plane, angling your racket face just a couple degrees forward or back will create topspin or slice, respectively; keeping it perpendicular will produce a flat, spinless drive. Horizontally, adjusting the racket face ever so slightly to the left or right, and hitting the ball maybe a millisecond early or late, will result in a cross-court versus down-the-line return. […] There are also the issues of how close you're allowing the ball to get to your body, what grip you're using, the extent to which your knees are bent and/or weight's moving forward, and whether you're able simultaneously to watch the ball and to see what your opponent's doing after he serves. These all matter, too. Plus there's the fact that you're not putting a static object into motion here but rather reversing the flight and (to a varying extent) spin of a projectile coming toward you — coming, in the case of pro tennis, at speeds that make conscious thought impossible. Mario Ancic's first serve, for instance, often comes in around 130 m.p.h. Since it's 78 feet from Ancic's baseline to yours, that means it takes 0.41 seconds for his serve to reach you. This is less than the time it takes to blink quickly, twice. (Wallace, 2006)

The dazzlingly myriad different variables could not possibly be all calculated, evaluated and put into action if such processing requires the standard view of perceptual processing. There has to be some way of predicting the incoming input, to know *how* to act before you are ever consciously aware of what you are acting *on*. This implicit knowledge may take many forms, and within enactivism there is much talk of 'affordances' and 'contingencies' (O'Regan & Noë, 2001b). They all take the view that they are ways of improving how an agent acts within an environment without necessarily encoding any rich information about the environment itself. A simplistic caricature would maybe be a series of reflex movements all allowing for further perceptual sampling of the environment until a goal is reached (although this is not one to utilise when supporting enactivism!).

Critics of the enactivist view who can see beyond the representational elements argue that this view dissolves the border between agent and environment too much (Kyselo, 2014). There is no clear dividing line between the two, and thus cognition bleeds into the world as well as receding depending on the situation. Some have even suggested that enactivism removes these boundaries entirely, and cannot avoid finding itself in a position where an organism lacks any meaningful border whatsoever[1].

---

[1] This is a conflation of enactivism with the extended mind hypothesis popularised by Clark and Chalmers (1998). Though enactivism leads nicely into thinking in an extended mind fashion, it is not committed to holding this view.

The boundaries between organism and environment are still defined however. In fact they are much better defined by this position. Rather than having an inner/outer dichotomy externally realised from the organism and the world, the two interacting is what generates a clearly delineated boundary. It is key here to note that the interaction is important *because* the organism and environment are constitutive of each other. The interaction is what creates these boundaries, and thus what creates the entities 'organism' and 'environment'. There is no way of separating the two into individual, independent elements without losing the conception of them entirely.

The classic example of this is that of a single bacterium swimming up a sucrose gradient (Brito & Marques, 2016; Thompson, 2005). Molecules of sucrose only possess meaning (i.e. that they are food) when they are in the presence of an autopoietic system which requires it for its continued maintenance (the bacterium). The sucrose only acquires meaning because of how the bacterium acts in relation to it. More importantly, this meaning arises from the action of an autopoietic entity, thus removing the need for an external observer to ascribe these meanings to what occurs. Without the bacterium, the sucrose has no intrinsic meaning of 'food'- this is a relational feature arising out of the usage of sucrose by the bacterium[2].

Thus, the idea of cognition leaking from out of the skull is also not quite as radical as the idea first seems. In a rather simple way, the writing out of your thoughts onto a page can be seen as a move of some parts of your cognitive processing into the external world. This is not to say the words themselves are cognitive, but that the cognitive process has become extended such that external elements to the nervous system are at play in retaining information (which only exists if an autopoietic system is present).

A useful analogy to bring in at this point is that of the Internet. In theory, we have access to all of human knowledge up to this date. We possess the capacity to access it virtually anytime and anyplace, and yet the amount of we actually access day to day is only a small fraction. Most of our daily interactions with the Internet consist of the same websites and similar content, and only rarely do we venture outside this to novel websites to find new information, say to complete a project or find a location nearby. Just because the sum of human knowledge is right there in front of us does not mean we are aware of it or access it, or even that we store

---

I myself find the two to be complimentary enough to combine well, despite the worries of certain critics. Answering why is not the aim of this paper however.

[2] The sucrose exists in the world as a chemical compound with physical properties whether the bacterium is there or not, but its status as food is only there if there is an autopoietic entity present with which to provide it this property

it for future use. Instead we rely on it being there as and when we require it, that we are able to find the content we need at a given time because we know *how* to look for it, even when we do not know what *exactly* we are looking *for*. All we retain personally is the means of accessing the Internet and some items we come across repeatedly every day, such as the URL of a website or an email address.

Replace "Internet" with "world", and the relation to enactivism becomes clearer. We know how to find the information we require in the world on the fly, we do not need to retain everything we encounter in a day to successfully negotiate our environment. Our cognitive processes provide us with the capacity to act in a way which achieves the correct sensory input to complete our goals.

Note in our analogy it would be rather odd to describe a URL as representing a website it allows access to. It most definitely indicates the website and the content it contains, but does not in of itself possess any of these contents. It merely tells us where to look. In the real world, our 'URLs' are the cell populations which suggest where to look, or what to move or to listen to to find the necessary information. They indicate to us areas of interest without representing the interest itself.

When you begin to look, there are many examples in both humans and animals of leaving the information to the environment around, to be accessed as and when. Territories are marked with scent, not retained as specific locations in the head. When we need to learn about tax practices in the Ottoman Empire, we look it up in a book.

The subtle point to suggest here is that our brain is not initially geared toward retaining information in any detail. Language has had the boon of allowing us this add-on, whilst cursing researchers to fall repeatedly into assuming or concluding memory and perception's primary functionality is representative and contentful. That we can learn of archaic tax codes and retain that internally is as a result of a separately evolved capacity to do so. Its roots however lie within a system that concerns itself with how to get knowledge as and when needed to further our ability to achieve our goals.

There are a number of different flavours of enactivist thinking, of which I intend to use autopoietic enactivism as my basis for the defining of declarative memory in a new way. To justify this however I will first explain each in turn, explaining why the other two do not quite fit within the argument that follows.

I.I.iv Autopoietic Enactivism

The notion of autopoiesis was first laid down in work by Humbert Maturana and Francisco Varela (Maturana & Varela, 1980). Autopoiesis defines life as both autonomous and self-reproducing. A system that is autopoietic is made up of elements which act to give rise to the system, as well as maintain the elements themselves (Weber & Varela, 2002). These elements also define the system as being distinct within the environment they occupy.

Autopoiesis is mechanistic, (usually) rejecting teleological methods of describing cognition (De Jesus, 2015; Di Paolo, 2005). The notion of autonomy (or agency), that an autopoietic system is in control of its own behaviour, is a core facet of autopoietic enactivism (**AE**). It is not however regarded as mechanistic[3], and is inherently teleological. Those who accept **AE** as a paradigm see the nervous system as an autopoietic system, entailing it is responsible for its own continual maintenance (Villalobos & Ward, 2015).

Within autopoietic systems, the idea of 'input', 'processing' and 'output', or 'start' and 'end' are definitions attached only by an external observer of the system. The processes themselves possess no such definitions of discrete steps in the ongoing dynamics of the system. Clearly, this shares much with Dynamic Systems Theory (**DST**)- representations and the like are not required to explain the functioning of the Watt Governor, for example (Van Gelder, 1995). **AE** looks at the spatio-temporal properties of cognition in a dynamical, non-representational manner which resembles other non-traditional paradigms of embodiment and embeddedness. We shall take **AE** as considering organisms to be *enactive agents*. The agency combined with autopoiesis allows them to generate their environments where areas of meaning and value exist only because of the presence of the organism itself.

Dennett (2011) argues that **AE** does not offer anything truly different to the genocentric form of evolution he supports. Yet this is to miss a key difference in the notion of evolution and natural selection within enactivism. When we speak of an organism and its environment in the traditional manner, the two are distinct. The inner workings of the organism are controlled for by the independent environment it exists in. The two interact, but both exist independent of the other. The organism is subject to the environmental changes of the external world and can do little to impinge upon this. Emergent, random mutations of DNA can provide beneficial attributes which enable an organism to best fit the environment, and thus more

---

[3] I must give room to begin to clarify this point here. Autopoiesis as Maturana sees it is one that holds mechanistic views in the more traditional sense. Enactivist views are not mechanistic in this sense, but are in a wider sense that they act to find commonalities for cognitive aspects whilst still being aware of the organisms' relation to the world around it on an individual basis. I'll unpack this further when exploring linkages between autopoiesis and memory later on in the text (Section III.II).

likely to reproduce. This information is resident within the genes of each cell, and is passed on to subsequent generations as a beneficial trait which over time may be adapted further to better suit the environment.

Enactivism does not share this view (Thompson, 2011). The organism and the environment are co-emergent. One does not exist without the other. The environment of an organism is entirely dependent on those external features the organism exhibits directedness towards and ascribes importance to. Two organisms within the same physical area will possess distinct environments as a result of the activity each performs and the actions of other organisms around them. Thus our genes do not adapt to the environment as much as generate a set of condition to adapt *to*. This is not to make the fallacy that a change occurs in the environment, *and then* we adapt. We are as dynamic agents constantly updating and adapting to our environment, which in turn is constrained by our own actions. This environment is what we then adapt to, not some objective, independent world.

Autopoiesis and **AE** are not completely in agreement, with some highlighting substantial divides in their conclusions about living creatures (Villalobos & Ward, 2015). Most of this appears to be the conflating of autopoiesis as originally envisioned with the work of Hans Jonas. His work focussed more on the organism, providing it with a privileged identity within the interaction of itself and the environment it is a dynamic relationship with. We shall look more carefully at this later, but in short I maintain the usefulness of the **AE** position over 'pure' autopoiesis, even with its potential shortcomings.

I.I.v Sensorimotor & Radical Enactivism

Sensorimotor Enactivism (**SE**) makes a provocative and persuasive claim as to how we engage perceptually with the world. When we look to act in the world, we do not do so because of any careful perceptual processing of the scene beforehand. Instead, we possess the capacity, or the contingencies, to act within the world. In knowing how to act, you sample the environment to achieve goals without ever having to transfer that environment into an internal one to be manipulated before action (O'Regan & Noë, 2001a).

This is the idea that the world does not require any representation because we can sample it as and when necessary for our current activity. A missile can hit a target, moving and avoiding obstacles, following the object's trajectory, without ever having any need to perceive this before acting. Rapid movements and actions are so fluid it is difficult to presume that we have in that time taken the whole scene in prior to our motions (O'Regan & Noë, 2001b).

SE suggest that the vast majority of our access to the world is geared towards action in order to perceive, and not the other way round. The more we interact with the world the more we master these contingencies and become more optimal. Whilst a well-supported area of enactivism, SE appeals too much to procedural memory and reflex actions to be best suited to our explanation of declarative memory. Sensorimotor research will no doubt benefit from adoption of some of the insights of this way of thinking.

Radical Enactivism (**RE**) holds a position which is difficult to grasp initially. The claim 'basic minds without content' is used as both a rallying cry (Hutto & Myin, 2012) and a denouncer (Miłkowski, 2015), and tends to result in masking the importance of the questions **RE** raises. What is being claimed is not so much that certain organisms are mere automatons with a stimulus-response relationship with their environment. Instead (and is made clearer in Hutto 2007), the claim is made to firmly place representations at only the highest level of cognitive processing.

At lower levels, the problem is that despite the decades of debate on this topic, none of the researchers in the field are able to settle on any universally accepted definition of even what a representation is. Because of this, each component of representative thinking depends upon the existence of another, which in turn depends on another. This final component depends upon the first, and leaves us back where we began. **RE**'s main argument is that the individual notions of intention, content and representation fall into this circular trap (Hutto, 2007). Hutto and Myin feel there is no adequate naturalised account of these three (particularly representation). As none are able to be held without the others, and none have an adequate explanation, they should be rejected as explanatory devices unless completely necessary.

To extricate oneself, you could resort to some strict and rigid definitions and enforce this upon the field. More easily though is to just do away with the notion of things such as 'content' entirely. In taking a position of interaction, and blurring of the traditional input/output and inner/outer relations, the need for internally maintained content goes away[4].

This is all to get ahead of ourselves. We will return to representations and the meanings of the terminology outlined above shortly. What is necessary to remember is that when we speak of

---

[4] I wish to make one final point on why we are focussing on **AE** rather than **SE** or **RE**. Unlike the two other main flavours of enactivism, **AE** is committed to understanding cognition as a whole. The other two focus on perception and basic content respectively; whilst key parts of understanding cognition, they are too limited in scope to grasp how we wish to approach memory.

enactivism rejecting representation, it is not always a *complete* rejection, but more a plea to actually look at the cognition going on, as opposed to how we should best categorise it.

I.II Representational Warfare

I.II.i Symbolic vs Subsymbolic

The level at which representations manifest themselves as we have seen are central to all three major positions we have mentioned. The most readily understandable is symbolic processing, where our cognition acts on types and tokens which stand-in for a given feature of the environment (Newell & Simon, 1976). This is akin (in a rather coarse manner) to a work of art or allegory, where the actual object is there to express something else. Taking our use of navigation in this paper, the external features of the world are internally realised as a representation, and it is this that cognition acts upon.

Connectionism has found this too hard to stomach, and an artefact of our own way of thinking which does not point toward what the actual mechanisms that underlie it do. But features such as line, colour, and shape are all categories that seem to arise naturally in the visual systems of all species we can test. Therefore there has to be some commitment to representing in a manner which allows for symbolic attribution after processing. In a connectionist network, each node in the middle layers receive inputs which are a blend of the features of the input. Thus any symbol or category is spread, or distributed, across these hidden layers, and combined only at the output level into features. Thus the symbols are not discrete, and operate on a level beneath normal conscious thought. These are referred to as subsymbolic for this reason (Smolensky, 1988).

A large number of relatively simple experiments have shown just how powerful this kind of processing is at recreating complex mental activity and computation, including face recognition (Burton, Bruce & Hancock, 1999), and grammar production (e.g. Plunkett & Marchman, 1993; Christiansen and Chater, 1999). Of course, there are problems with this setup as well. These networks are restricted by the inability to self-generate input, and are often highly suited to the training data they have had with little adaptability to other tasks. In addition, many require back-propagation of outputs to train the network. This can be after being informed that the output was correct by an external 'critic', or by internal comparison to the input. But each of these has little biological plausibility in many situations that arise in day to day living. And whilst there is some evidence for back-propagation (Stuart *et al.*, 1997; Waters, Schaefer & Sakmann, 2005), there is not a sufficiently coherent amount to offer this as a primary way of learning.

One way of avoiding this practicality issue is to ascribe correcting powers to networks higher up in the processing chain of the brain, or those within the same level checking nearby processing. This idea is at the heart of predictive processing accounts of cognition (e.g. Clark, 2013, 2015), and are complementary to some of the ideas which will be sketched within this paper.

I.II.ii Indication is not Representation

The oil change light on the dashboard of a car does not represent that the oil requires changing. It indicates that something has occurred, but not in any contentful manner. Similarly, the words on a page such as this indicate the meaning of what my current thought process is, without itself representing it. To represent requires an observer, a means of assigning an action, object or process the property of standing-in for another.

Indication is analogous to coincidence detectors, oft talked about with regard to activity at the synapse. **NMDA** receptors, a type of voltage-gated ion channel critical in memory, are blocked by a magnesium ion in the pore. This ion is only dislodged by a combination of a conformational change caused by the endogenous **NMDA** agonist glutamate, and the depolarisation of the membrane. Both together indicate that the cell has received sufficient input from axonal projections to elicit an action potential. **NMDA** receptor opening allows a greater sodium influx into the cell and also initiates calcium signalling. It is this which brings about mid to long term biochemical changes that result in learning.

In *Representation Reconsidered*, William Ramsey (2007) presents an argument that emphasises the above example as not being representational. The receptors are responding to a particular stimulus, and as they carry information on it, they must be representational. What Ramsey claims however is that defining this as representational does not provide any more usefulness to our understanding than referring to them as causally linked to a given stimulus. The explanatory power is the same. To claim representational status is to make a strong claim about this relationship between receptor and object, and one which may result in making the term 'representational' useless. If everything is deemed representational, the term loses any real meaning and results in conflicting ideas sharing the same terminology.

This position tries to provide a 'job description' to representations, one which is robust enough to allow for their existence but only when they possess explanatory power that surpasses mere causal linkages. In recent papers, Ramsey has qualified his scepticism over the nature of representations (Ramsey, 2014, 2016). You can commit to representationalism, but

must not use representations to define cognition itself. These definitions often stem from our 'folk' notions of cognition, and do not stand up to scientific scrutiny (Ramsey, 2014) Non-representational states are likely to exist in the brain, and appealing to them should be a part of our explanation of cognition. We should avoid making too much of these theoretical notions however:

> Cognitive science should not to [sic] be defined by the *type* of explanation being offered. Cognitive science should be defined by the *target* of the explanation. It should be demarcated by the kinds of questions we are trying to answer, not by the sort of answer that is on offer. (Ramsey, 2014, p.12)

I.II.iii Neuroscience & Representations

Unfortunately for our question, representations are very much entrenched into the very framework of neuroscientific investigation (henceforth neuroreps). When empirical data indicates a cell fires reliably with a specific behaviour or sensory stimulation, it is intuitive to assume this is a causal relationship. Given feature x, cell y fires because of x being present. Subsequent work aims to support or refute this claim, but rarely ever doubts the relationship as being causal. Neuroreps are present from the very outset of planning and theorising within the field. Bechtel (2016) illustrates that this underpinning of the work precludes any attempt to dismiss it as a descriptive 'gloss'.

Ramsey (2016) has highlighted that the neurorep does not map neatly onto the definitions currently found within philosophy of mind (henceforth mindreps). Indeed, he dubs it a 'hybrid stance'. Neurons represent because they fire in co-variance with some environmental cue at a reliable rate. Content is presumed to be a part of representations as a first assumption. Subsequent analysis aims to match this to the vehicles of said content, be that a neuron, a protein or a network.

In this way the empirical work, that which assumes all this to be true, provides a mechanistic explanation within which content/vehicle relations are an integral part. Representations are not, as some assume, a 'gloss' that is used to explain the data. They are extant from the very start. Therefore, we cannot take any empirical data and claim it does not point to a representative explanation. Yet this is not a death knell for any non-representational approach. We can accept this neurorep commitment, take that the work assumes their existence. What is important is that we then show that such representations are performing a *descriptive*, and not an explanatory role. If they act to help describe the empirical work, provide a handle on the sub-personal workings of the systems, then they do not interfere with the notion that the activity itself is not representational.

These representations are far removed from the philosophical notions of content, vehicle, intentionality and the like. What they do commit to is that the internal neural firing covaries with a configuration out there in the world. Thus these neurons which fire carry information about what is occurring in a manner easiest to presume as representative. Neuroscience then studies this covarying in an objective manner.

I.III A World Enacted, Embodied and Embedded

Rejection of representations is often linked to a rejection of the notion that when you see the world, you take it all in. Any failure in returning this experience accurately is due to incorrect maintenance or access to this information. Yet such a veridical input and the necessity to retain it would be a heavy burden on the system. If we weaken the view, take that the sensory inputs take in as much as possible, and we represent to enable our filling the blanks, then we are still in a quandary. This kind of internal processing of a degraded input only makes sense if the cognitive agent has suddenly been removed from the environment and placed elsewhere. Otherwise, the agent has the capacity for continuous sensory data inputs to fill in any blanks as and when the system requires it. But once again we run into an issue. We would not say that our experience of the world, rich as it appears, involves repeatedly filling-in and representing only bits at a time what we experience.

The enacted, extended, embodied and embedded (**4E**s) approach to cognition is at its core a push towards revising our textbook and knee-jerk notions of how thought must occur. The representation issue is really just an easy place to start, given its already shaky foundations. You can accept representations in all **4E**s, but they are not seen as intrinsic to thought such that they exist in every cognitive agent on the planet. Doing so can lead us dangerously close to 'homuncularising' thought by finding evidence for proto-forms of cognitive processes we assume to be reliant on internal representations.

Instead, the answer is to maintain that our ability to represent can only be reliably attributed to ourselves. This ability is reliant on the blending of perception and action processes (as these positions attest to being more intertwined than the traditional stance), which then allow for it to emerge. But we cannot explain cognition using internal representation because this would mean we explain our end point (representational capacity) with the very thing we are trying to explain.

When continuing this discussion further, I shall refrain too much from getting bogged down in intricacies of why and how we should say any particular thing is representational or not. It is sufficient to say that any empirical work presented will in some way assume the existence of representations, and the enactive positions will not. I will leave it up to someone with more time (and patience) to resolve that tension. For now, we need to move on to our cognitive process of choice: memory.

I.IV Memory

Despite centuries of work, we are still seemingly unable to make any real grasp on what *exactly* we mean by 'memory'. More often than not we break it down into various types and forms (as we will do below), and our very basic descriptions appears to change readily with technological advances as well as cultural shifts. Arguments of whether particular memory types can be defined as natural kinds can be found for each major branch (e.g. Michaelian, 2010, 2015; Werning & Cheng, 2014).

If pushed to find a general definition, one may say that it is a set of cognitive systems which are loosely connected by their desire to retain information from the past to be used in the present through reconstruction or action (Sutton, 2015 has a similar definition). Commonly, this means the maintenance of some content within the organism that is stored away until needed. There is an input, which is encoded into some physical form, which is then consolidated and maintained. Upon retrieval, this information is unpackaged from wherever it is being maintained, and then acted upon by the organism. It from this definition that we springboard from.

Here I adopt the terminology used within neuroscience. The most used way of broadly splitting memory is into declarative and procedural (non-declarative) memory (Squire, 1992). Riding a bike, knowing how to write and being able to play the piano are all examples of procedural memories. They all share the property of being motor skills; it is our capacity of knowing which actions are required to perform a particular complex motion. Given this, it is hard to verbalise these memories, as they lack any easily afforded description that does not slip into our second type of memory.

I.IV.i Episodic Memory, Semantic Memory

The focus of our jaunt into memory work is the type we utilise every day. Recalling where your keys are, the conversation you had with your friend yesterday, all are examples of episodic memory (**EM**). Succinctly, **EM**s are those which pertain to events and locations, even

if those have not occurred to you personally (which fall into the further subset of autobiographical memories). Semantic memories are those of 'facts'- memory stripped of context to exist on their own as some form of entity. Such examples of this are that Luxembourg is a landlocked country, or that apples can be red or green.

Each has a long history of research behind it, which we will visit throughout this reconfiguring of our conception of these memory types. We start with the most famous research program, the hunt for the neural correlate of memory.

I.IV.ii Learning and Engrams

The "search for the engram" has been the preoccupation of memory researchers for the most of the last century (Josselyn, Köhler & Frankland, 2015). The engram search is obviously a representative notion- what changes to the neurophysiological makeup of the brain constitutes a storage unit of memory? For a long time, this engram was never thought to be contained even in one brain region, let alone single cells.

Lashley (1950) trained rats to navigate mazes and subsequently lesioned various parts of their brains. He noted that the more brain area was lesioned, the greater the subsequent impairment on the maze. What was *not* important however was the area of brain removed. Lashley's experiments led to rejection of localised functions for a time, but the explosion in new techniques for analysing the brain has shown that regions can have very specific functions even if large portions of the brain are involved. There is also a lot of redundancy in the whole system (Friston & Price, 2003). The focus now is back to engram networks and cell types holding specific jobs within the brain.

Memory engrams have been the pursuit of many labs, with some claiming to find evidence for specific memories encoded in a small subset of cells in the brain (e.g. Roy *et al.*, 2016). Yet there is the acceptance that long-term memories may never be fully consolidated (Dudai, 2002). Simultaneously we have a system which appears to require specific brain regions to form memories, but does not require this same region when recalling a memory held for a long time. To explain such a scenario, we must turn to the functional aspect of the hippocampus and how we understand short and long-term memory.

I.IV.iii Hippocampus and Entorhinal Cortex

The **EM** 'wetware' has grown to encompass more than just the hippocampus. In recent years, each anatomically distinct part proximal to it has been painstakingly studied and categorised

according to the results of lesion and pharmacological studies. For our purposes, we need only focus on two areas, the entorhinal cortex (**EC**) and the hippocampus.

Entorhinal Cortex Split into medial (**MEC**) and lateral (**LEC**) parts, the **EC** has been the main focus of spatial cell research outwith place cells. A substantial population of cells in the **MEC** fire in a crystalline pattern with fixed spacing, and have been dubbed 'grid cells'. This system appears to play a major part in path integration, which is the ability to navigate through space only relying upon egocentric metrics. We shall go more into this later, and for now will focus on the **EC** more generally. The **EC** conveys information from associative areas of the brain to the hippocampus. **MEC** supplies spatial information, whilst **LEC** is more concerned with objects and their relations within the environment (Eichenbaum *et al.,* 2012). Studies have placed the **MEC** as necessary for both hippocampally-modulated spatial navigation (Schlesiger *et al.,* 2015) and memory formation (Kitamura *et al.,* 2014). For us, what is important is that most evidence points to it playing a major role in both memories and navigation.

Hippocampus This almond shaped region of the brain has long been associated with memory. Bilateral lesioning of the hippocampus results in an inability to maintain new memories, whilst leaving the rest of cognition relatively intact. It possess robust connections to many other regions of the brain (Lavenex & Amaral, 2000), and has been linked to involvement in emotion, navigation, learning and memory. Only declarative memories require the hippocampus, with procedural learning unaffected in lesion studies. The extent to which this region is necessary after encoding of the memory is debated within neuroscience. One field of thought contends that all **EM**s are encoded and maintained for a time in the hippocampus, before being consolidated elsewhere in a distributed manner (Wang & Morris, 2010). The other agrees with this for semantic memory, but maintains that for **EM**s the hippocampus is still required to act as an indicator of where these memories are consolidated, much like an index card system (Nadel *et al.,* 2007). Recent work points more towards the former, with memories for familiar events only requiring a short time within hippocampal networks before consolidation elsewhere (Tse *et al.*, 2007; van Kesteren *et al.*, 2012).

I.IV.iv Neocortical storage

If the hippocampal system is responsible for initial processing and short to mid-term memory, where do our longer term memories go? The answer is everywhere. That is, the network states for a given bit of information deemed important are distributed across the whole neocortex (Wang & Morris, 2010). Going back to Lashley and his rats, such information is retained precisely because of this redundancy in the system.

By disseminating the recall as a global neocortical state the memory is relatively secured from minor neurophysiological trauma and plasticity at the synapse. To do so in an efficient manner much of the context of certain memories is stripped away. What is left is some key component of an **EM** which when reactivated may also reactivate a state where the context re-emerges. What is important here is that this long term memory is not stored in the same way as books in a depository. The memory itself does not exist in isolation, and if it possesses a physical correlate it is likely to share these with others. Instead, a memory of this type is active when the system returns to a state which re-fires a sufficient number of neurons in the correct dynamical pattern. Simply put, a memory does not exist until the system is placed into the same or similar dynamical pattern in which a previous state occurred.

At this point most readers will rightly point out a flaw here. If we must return to the environment in which a memory occurred, how can we recall things which happened many miles away in places which may not even exist anymore? Surely here there has to be some storage of the relevant details to be able to recall? Whilst an argument may be made for this, remember we previously state a memory does not exist in isolation. Our faculties are constantly ongoing, with memories of many kinds constantly being brought into being. When we wish to recall what we did on our last holiday or the plot of a book read long ago we can ride these states to the memory we desire through association. Patterns of dynamics are what are important, over the need for the exact same environment.

Humans have the unique capacity to jolt the brain into specific states by means of language. Language likely acts as an index card system for previously realised brain states, constructing an environmental scaffold upon which memories can be regenerated (Carruthers, 2002; Marian & Neisser, 2000). This tool has been immensely useful in augmenting our capacity to plan, act and communicate. It works so well that we often forget that much of the underlying processes of memory do not require language at all. The functionality of our declarative system, however you think of it to be, is not dependent on language. It benefits from its involvement, but it plays no constitutive role in what goes on when we say a memory is 'stored'.

I.IV.v Oscillations and Information Transfer

How do **EM**s become semantic ones? The current theory is that the information from the hippocampus is transferred by way of oscillatory phase locking during periods of rest and sleep.

This information is transferred from one location to another via sharp wave ripples (**SWR**s)[5]. These are transient bursts of 110 - 200 Hz originating in the **CA1** region of the hippocampus.

These ripples act as information packets, sending the day's collected system states into the neocortex where they can be attached to relevant schemas. Exactly how they do so is unclear. After all, if these ripples send information in a scrambled manner, there is a need to decode this somehow on the other side (Our dreams may be a perceived consequence of this process, although this is highly speculative).

SWRs are sent during moments where the oscillatory pattern of the hippocampus matches with that of the medial **PFC**. This locking of the phase syncs the two regions, and thus allows for the state of one to impinge on the other. To avoid catastrophic interference, such locking events occur for only fractional time windows. The two regions are in and out of phase many times each second. Disruption of these ripples during sleep is sufficient to weaken memories of the previous day to the point of forgetting.

As the above description attests, the information-theoretic notions are the best way of explaining these phenomena, although the appearance of network states points towards a less discrete symbolic processing. The tricky task of "enactivising" this we shall tackle later in **Section III**. For now we must go down a level to the cellular basis of our memory mechanism.

I.V Of Mice and Men (and Rats)

I.V.i Cognitive Maps

The idea of a cognitive map first arose in 1948. Edward Tolman, in a series of elegant experiments, appeared to show the brains of rats were capable of representing the external space around them. Such a representation was proposed to take the form of a 'map', as rats made shortcuts which were not trained in order to reach their goals quicker.

The hippocampus appears to be the core unit of this cognitive map system. Functioning as it does as a region of sensory integration, this combining of spatial with non-spatial indicated a potentially good fit with the idea of internal maps of the environment (Manns & Eichenbaum, 2006). Although not without its detractors (e.g. Bennett 1996), the cognitive map hypothesis (**CM**) has been robustly supported by evidence and theory for the past half century. The first

---

[5] This section takes its information from Buszaki's exhaustive review (2015) of sharp wave ripple research, which is comprehensive in the literal sense of the word.

major piece of critically neural evidence came from the discovery of cells that only fired when a rat was in a particular area of an environment, facing in a specific direction (O'Keefe & Dostrovsky, 1971).

Since this discovery, there has been many more cell populations that exhibit firing properties supporting the notion of an internal topological map. The complexity and encoding capabilities of these cells has led some to consider the chance that they also play major roles in encoding and consolidating **EM**s. In order to make sense of this claim however, we must first introduce these cell types and explain why they are seen to be representative of the world around them.[6]

I.V.ii CM Cell Populations

Place Cells These cells encode information about a specific location (or locations) in an environment. Within a given environment E, the firing profile of a place cell will not alter as long as there are no major changes to E. Achieving this seems to require the integrative abilities ascribed to the hippocampus of both non-spatial and spatial data (Marozzi & Jeffery, 2012), and these cells are mainly situated in the **CA1**, **CA3** and dentate gyrus regions of the hippocampus (Lee *et al.,* 2004).

Place fields show a Gaussian firing property, with greatest firing at the centre of the field, becoming progressively lower as the organism moves towards the edge (Burgess & O'Keefe, 1996). Within a population of place fields, all which overlap with others at least partly, this provides accurate data of the organism's current perceived position within E. This is best illustrated when you navigate in the dark. If placed in E and allowed to explore before being blindfolded it is likely you'd be able to navigate around quite successfully. This is thanks to this population of place cells and their overlapping fields. It is harder to be uncertain of your location when the firing of other cells provides reciprocating input to reinforce or dampen firing of a given place cell. Yet in a situation such as this you are likely to not be as fluid in your motions or your accuracy of place as when visual feedback is available. This is a general feature of our navigational system, in that self-motion and prior experience are sufficient for navigation, but lose accuracy as time since the last perceptual update has occurred (Zhang *et al.,* 2014)[7]. Much like many of the other cell types listed here, the topological mapping of the space requires an initial exploration of the environment. If you were to be blindfolded and placed into a house you had never been in, you'd find it almost impossible to navigate successfully without considerable trial and error. Activity in E is what entrains the correct

---

[6] For a more comprehensive review, see Moser, Rowland & Moser, 2015

[7] A nice summary of what memory may also turn out to be.

patterns of activity to navigate successfully around it. Self-motion cues can only do so much, we rely on our sense of sight, smell, touch etc. to speed up and refine the process.

Although stable firing fields are a property of place cells, this does not mean they are invariant. Cells 'remap' due to global gross changes to E, changing their firing field(s) completely within E as if it was a new room. If the place is reverted back to the old configuration, so does the firing fields of the place cells (e.g. Knierim *et al.,* 1995). In addition, these cells may exhibit an alteration in their firing properties whilst the actual firing location remains constant (Leutgeb *et al.,* 2005). This 'rate remapping' indicates the encoding of non-spatial information in place cell activity, as the phenomenon occurs in conjunction with changes to colour, shape (Leutgeb *et al.,* 2005), odour (Anderson & Jeffery, 2003) etc. How these are brought into the system is not fully understood, with grid cell modulation as a primary input (Solstad, Moser & Einevoll, 2006) recently being called into question (see Bush, Barry & Burgess, 2014).

Temporal dynamics are encoded through use of the natural oscillatory rhythm of the hippocampus. Theta oscillations (7 - 10 Hz) occur only when an organism is moving (Lisman & Redish, 2009). The firing field of a place cell is approximately 1 second across, or 8 cycles of theta. The firing of the place cell occurs every so slightly earlier within each theta cycle as the organism moves through the field (O'Keefe & Recce, 1993). This phase precession generates a temporal code. For a given population of place cells, each region in space will have a specific coding profile dependent on these phase precessions. Thus, if one can read out these firing patterns, you can subsequently reconstruct where the animal is, from where, and where it is likely to be heading towards. This gives the place cell population the capacity to calculate trajectories through space in a flexible manner.

Grid Cells Upon entering an environment, the space is mapped out by the regular firing fields of this cell type. These cells fire in triangular arrays spaced equal distances apart within a given population (Hafting *et al.,* 2005). This spacing is fixed and changes by specific amounts across the dorsoventral axis of the medial entorhinal cortex (**MEC**). This creates a series of grids from different cell populations which collectively map the entire space at all scales necessary for the organism (Stensola *et al.,* 2012).

The grid cell is a slight misnomer, with two different distinct cell types being able to classify themselves as such. The initial assumption was that pyramidal cells in the **MEC** were the grid cells, surrounded by stellate cells more specialised to borders (Tang *et al.,* 2014). Yet recent work has questioned this, finding cells within both cell types specialised more for speed, head

direction, or both grids and head direction (conjunctive cells) (Sargolini *et al.*, 2006; Zhang *et al.*, 2013).

Head Direction Cells Firing of these cells is tied to a Gaussian distribution around a preferred direction anchored to the animal's' head position (Evans *et al.*, 2016). All head direction cells (located in the **MEC**, thalamus and postsubiculum) have primary projections into the **EC**, combining their inputs into the grid and place cell systems (Taube, 2007). Much like the other cells mentioned here, their firing is stable for a given environment.

Border Cells Boundaries are key features of any environment and are seemingly encoded too within the navigational system. These cells provide input to the place and grid systems on the layout of the surroundings, and are vital to the stability of these cells firing fields (Solstad *et al.*, 2008). The longer a boundary is not found, or the firing of these cells inhibited, the more drift that occurs of place cell firing fields (Zhang *et al.*, 2014).

Speed Cells Localised to the dorsal **MEC**, the running speed of the organism modulates these cell's firing rates (Kropff *et al.*, 2015). The greater the speed, the greater the firing rate, with so far nothing to show anything other than a direct linear relationship robust to any changes elsewhere. The existence of these cells has been implicated in the updating of grid cells, as well as the generation of these cell types in general.

The spatiotemporal information seemingly encoded by these cells places them in a position to provide the bedrock of any **EM** system (Marozzi & Jeffery, 2012). The minimal requirements of declarative memory are that it pertains to a place, in a time, and that events unfold directed to and by things that resemble previous perceptual states. Given that the hippocampus is undoubtedly the critical hub of **EM** processing (Tse *et al.*, 2007), this assumption is on strong ground.

If we frame the above in the more traditional notions of cognitive structure then this link to memory is obvious and uncontroversial. Place cells, grid cells and speed cells, if we take the traditional view, provide a highly detailed internal representation of the surrounding environment, a cognitive map. This richly informative network can then be acted upon not only to act, but to provide a snapshot of what occurred. Memory is then just the recycling of representations generated whilst moving through the world, with repeated exposure resulting in long term structural changes which 'fix' these representations within the brain.

I argue that whilst a plausible interpretation of the data available, it forgets that such a system appears to rely quite heavily on an organism's activity being an ongoing process. Immobilising animals and placing them within an environment results in no acquisition of stable place fields (Foster, Castro & McNaughton, 1989). Whilst it is true that the system still co-varies with the environment, and carries information in a representative manner, this is not sufficient to say this is a representation in the philosophical sense. The cells fire as a result of organismal activity, and are just as much involved in laying down the environment we perceive sensorily as it is in mapping it out. The system provides a good *indication* of what the environment is currently like. These indications are what feed into our memory, not representations.

Growing evidence from all these studies is that each cell type is both *robust* and *flexible*- exactly the two terms used by Thompson (2007) for minimally autopoietic systems. There are many parallel inputs, each sufficient to be utilised to form firing activity which covaries with place, area, boundaries etc. The navigational system is its own powerful autopoietic system, one that, as we shall see, brings forth memory as a natural result of its complexity.

Using the above knowledge, I will lay out the claim that how we intuitively perceive memory is a false construct that puts too much emphasis on our ability to interpret it over its actual functionality. Even when we research memory and understand these functions, I argue our current holding onto representationally dependent paradigms constrains our ability to understand our capacity to mentally time travel.

I.VI Rethinking Memory

I.VI.i Not Just for People

A common and quite easy mistake to make in assessing the cognitive properties of memory is to maintain a kind of anthropocentricism. This comes in two flavours. The strong one is that we, as humans, are in possession of a memory unique amongst animals. Our capacity for memory so exceeds that of other organisms we are forced to conclude any work in animal memory research will only act to add reinforcement of equivalent human work. I will not be so bold as to suggest this is consciously held by anyone seriously, but there does some to be some implicit sense in which this holds when speaking of memory.

Then there is the weaker version, where people are happy to accept memory is not particularly more complex in humans compared to other species, but still take such work to be 'lesser', or in need of comprehensive human work to be deemed 'reliably true'. It is an obviously trivial claim to make that if you want to understand human cognition, you should do so by using

human empirical work. But this is to ignore the fact that the vast majority of our understanding of memory comes from experiments pretty much impossible to perform in humans. The hippocampal formation is deep within the brain, and any direct activation would require extremely invasive brain surgery (not to mention be ethically unsound). We have gleaned much from natural lesions such as patient **H.M.** (Squire, 2009), yet none of this work allows for the direct manipulation required to evaluate new hypotheses with the old.

If we are to provide an account of memory which includes the history, the social and cultural setting (or ecological niche), and the actual neurophysiological structure of the animal, then we have to do away with this anthropocentric attitude. In the spirit of this, I intend to include very little in my account herein of memory of human research. The foundation of human memory is rooted in functions and structures shared by many organisms. If we are to build a new theory, it needs to be done from the ground up.

I.VI.ii Memory through the Ages

To see how memory has not always been perceived as it is now, we shall take some examples from Frances Yates' (1966) investigations into the history of memory from Ancient Greece through the Renaissance to the present day. The amorphous nature of the term memory has lent itself well to this shifting of identities and conceptions as a result of societal and cultural shifts, but it also highlights the active nature of memory as being more present than not.

The invention of the art of memory, or mnemonics, is attributed to Simonides, who was able to identify the bodies of people he had attended a banquet with after the roof had caved in, killing everyone inside. As Cicero notes, Simonides realised:

> […] that persons desiring to train this faculty (of memory) must select places and form mental images of the things they wish to remember and store those images in the places, so that the order of the places will preserve the order of the things, and the images of the things will denote the things themselves, and wc shall employ the places and images respectively as a wax writing-tablet and the letters written on it. (Yates, 1966, p.2)

Ancient civilizations held great stock in the ability to recite long passages from memory, and the technique of mnemonics was honed independently by many cultures. What is shared between them is two key things. The first is a separation of natural and artificial memory, and the second the requirement of using spatial references as a mnemonic device. These previous insights into memory-as-artform can provide us with a better grasp on just how declarative memory has evolved and how much of this natural/artificial divide may have been forgotten.

Natural memory, that which we all possess by default, is tied to thought as and when it occurs. This can then be reinforced or enhanced by the acquisition of an artificial memory. Here,

artificial memory is mnemonic memory. This system of learning makes use of memory spaces or palaces, in which items are deposited that stand in for the thing or things to be remembered. Often these are not one-to-one mappings, what is often more important is that the item appeals to more than just the visual sense.

A Roman teacher of rhetoric said as much in a passage detailing the necessary requirements of the images any memory space:

> […] things immediate to our eye or ear we commonly forget; incidents of our childhood we often remember best. […] nature shows that she is not aroused by the common ordinary event, but is moved by a new or striking occurrence. […]
>
> We ought, then, to set up images of a kind that can adhere longest in memory. And we shall do so if we establish similitudes as striking as possible; if we set up images that are not many or vague *but active* (*imagines agentes*); if we assign to them exceptional beauty or singular ugliness; if we ornament some of them, as with crowns or purple cloaks, so that the similitude may be more distinct to us; or if we somehow disfigure them, as by introducing one stained with blood or soiled with mud or smeared with red paint, so that its form is more striking, or by assigning certain comic effects to our images, for that, too, will ensure our remembering them more readily. The things we easily remember when they are real we likewise remember without difficulty when they are figments. But this will be essential— again and again to run over rapidly in the mind all the original places in order to refresh the images. (Yates, 1966, pp.9-10)

This one passage provides evidence that many of the underlying assumptions of an enactive memory are nothing new within mnemonics itself. Novel experiences enter memory more readily than the mundane, memories are stronger if tied to action and to be few and distinct, and repetition is the only way to "refresh" these internal visualisations. All three have a part to play in our resetting of how we approach memory as an object of inquiry, and all three now possess empirical evidence that supports them.

This training of our memories creates an understandable confusion when we consider that we seem to be appealing to stabilising some internal contentful state. Despite the action and perceptual saliency of mnemonics, the idea of memory palaces appears to commit to seeing memory as an archive, although this is not its intended purpose. It may be interpreted as a static store of images which awaits access, but this is only surface reading that occurs when unaware of the rhetorical past of such mnemonic devices (see e.g Hutto, 2016).

In our current understanding of memory, we neglect the fact that this art of memory is something we acquire through education, and that our own memory system is thus augmented by this procedure. Our natural memory system, the one which we wish to actually understand, is buried beneath this. Hence, looking at children and animal studies provide a way of viewing this natural memory removed from this training, and is how we shall approach the later sketch of an enactive memory. Education has profound effects on our way of thinking, masking

everyday forms of cognition by providing a scaffold for abstraction and generalisation away from the immediate environment (Luria, 1976). We possess memory systems which are aided by this artificial component, one which may well be representational and internally rich in the cognitivist sense. But this is not memory as it were, in the wild.

I.VI.iii From Navigation to Memory?

The evidence through time is that there has always been the implicit acceptance that memory requires some form of travel, and that it does not always maintain itself strictly within the confines of the skull. Take for example, attempting to recall earlier I mentioned tax practices in the Ottoman Empire. You will likely remember that I did, and that it was in a particular region of the page. But it is much less likely you will have anything more than a weak gist of what it was mentioned in relation to. Your recollection is dependent on the information of where it is, but not what it means exactly. For that, you have to search out the words again and regain the context.

Without the ability to write things down, or with illiteracy, this becomes more obviously apparent:

> The use of the place mnemonic is a natural outgrowth of the way memory for facts is keyed to our physical surroundings. For nonliterate societies, the places of the physical world become a sort of lived-in memory space. I stayed for a while in a remote village in Papua New Guinea, surrounded by rainforest. To my eyes, the forest was dense and beautiful, but without variety. Space had no meaning there. But my local companions and guides would often stop and point at some bent root, some certain small rise or gulley, and laugh, remembering something funny so-and-so had said in that spot, or remembering a particular pig they had killed there. The jungle was, for them, a familiar landscape of memory. (Wargo, 2010)

The ability to know *where* to return to in order to gain information is much less energetically costly than always knowing *what*. Integrating the environment into our memory system reduces the workload by allowing for access to states to be driven by stable external means. In our enactive sense, this is easy to explain through the organismal-environment.

As I alluded to earlier, our understanding of memory recently has been based heavily on spatial navigation tasks in rodents. The underlying assumption has always been that spatial memory *is* **EM**. To remember which heading to take, which directions and turns, and what cues to pay attention to, an animal must possess some capacity for remembering what has occurred previously.

This idea is not without detractors however. Some argue that **EM** requires a sense of understanding the events that you can recall are often uniquely your own, that they can be

autobiographical in nature. A rat solving a maze task does not have access to this self-reflection about the memories used to complete it. There is information present, and it is used by the brain to reach a successful conclusion, but this is never elevated to conscious thought. This episodic-like memory fails to reach the standards of human memory function (Cheng, Werning & Suddendorf, 2016; Suddendorf & Busby, 2003).

For this discussion, I take the position that knowledge on how to complete a task acquired by rodents during experimental testing amounts to **EM** in a way which enables its use in explaining the workings of human remembering. Even if the rat cannot self-reflect on the past it has experienced, the underlying mechanisms of the memory system itself remains the same. When in a familiar environment the features therein may act just like the outstretched roots or gullies of the Papua New Guinean tribe. Any additional self-reflection merely improves the ability to utilise this system for secondary purposes, such as remembering what we did.

I.VII Regaining Our Bearings

Thus far we have travelled on a somewhat breathless tour of our main themes. I have explained that in choosing **AE** as an explanatory paradigm for what follows I intend to largely reject the idea that memory is representational in the classic sense. The memory we are focussing on, declarative memory, is what determines who we are as it provides a history of us as a cognitive agent. This type of memory is also involved in the extraction of facts about the world, converted into more robust and abstract forms.

Work in the spatial navigation and memory of rodents has provided a wealth of data which shows very clear linkages between the two functions. Furthermore, historical conceptions of memory show clear commitments to splitting natural memory from the artificial memory made possible through language. These linkages run deeper than a surface similarity or shared mechanisms. Rather, they appear to be one and the same, with the only change one of complexity and distribution of this information. Memory research has been committed to representationalist views since the 1940s however, and thus much of the findings have only ever been assessed from this perspective.

Given the hypothesis that declarative memory arose out of ever more increasingly complex navigational neuronal systems, this paper's aim is to sketch out the idea that this history may mean our understanding of declarative memory requires an enactivist account. Such an account will be built largely from the work of Thompson and Varela, and will rely mostly on evidence gleaned from work in non-human organisms.

Given that there is some debate as to whether the neuroscientific data can be utilised in a manner which allows for a non-representational account, I shall also attempt to incorporate the more action-oriented memory hypothesis into cognitivist and connectionist frameworks. These will be compared to the enactive memory approach, and assessed by how well each can provide an explanation for declarative memory which does away with the idea of 'memory as remembrance'. Whilst I make some allusions to semantic memory (**SM**) and how this work naturally leads on to helping explain this side of declarative memory, space prevents me from developing this further. The focus instead will be on episodic modes of remembrance and thinking (although there will be space for tentative forays as we progress).

In this discussion, I hope to make clear that recurrent spatiotemporal patterns are deeply embedded within neural firing, and that it is this which links our accounts of action with so-called 'higher' cognitive processing. Our cognition is active and real-time. When we ascribe a primary function to a system, we do so through what *we* observe to be primary, which often favours the idea of deliberation and modelling of the world around us before making decisions or acting. This paper links up various researchers working to shift our perceptions away from this, and highlights the idea as being hidden in plain sight within recent empirical research.

II. Memory is Not about Remembering

II.I Misremembrance and Forgetting

I hope to show that declarative memory is not a system only functioning to remember past experiences. It is involved in predicting what happened, what is happening, and what *might* have happened (De Brigard, 2014a). This goes against our conventional idea of what our memory is there for- "The way in which a particular mental content is experienced by us is orthogonal to the purpose of the system that is responsible for providing us with such an experience" (2014, p167). Our perception of how memory works does not mean this is what it is trying to do.

Let us take instead the stance that our **EM** extracts rough ideas (gists) from experiences based upon an ever changing framework of related semantic information. This bank of knowledge is a generalised form of knowing how things relate in the environment based on previous interactions. After a few exposures to the same experience, there is little need to make any major alterations to what is already semantically held. What matters is the gist. Recalling your breakfast for the past seven days may be doable, but not each individual day's experience which

surrounded it. Ruling out a disruption to the normal way of things, the memory of the cereal or toast etc. are not really situated into a veridical recollection of each morning.

Despite our intuition that memory provides us with remembrances of things past, it does not necessarily follow this is its *function* (Allen & Fortin, 2013). It may be a secondary benefit, a by-product of a larger system. Current work on the aetiology of memory still focusses tightly on human remembrance, and neglects that this capability does not define memory itself.

De Brigard (2014a) showcases this point quite nicely. He suggests that the capacity to remember is part of a larger system whose function is to recombine previous experiences to bring about 'episodic hypothetical thinking'. Remembrance as we suppose it is based upon a 'content-based approach'- memory experiences are of the past, thus the memory system is there for this purpose. Occasions when we misremember or forget in this approach are failures of this capacity. Yet this is a strange thing when we consider just how often we forget, or fail to recall things accurately. A cognitive process which does so bad a job would have long been altered if this was how it was meant to function.

So let us adopt the opposite stance. Inaccurate memories, those with information omitted or altered, are not wrong memories, or an aberration in the system. They are the sign of a healthy one. Further evidence for this comes in two forms. The first is the memory capabilities of amnesiacs and Alzheimer's patients, who have major damage to their capacity for memory. The second is the case of Solomon Shereshevsky (more well-known as 'S'), who had the complete opposite problem: a memory which was seemingly infallible (Luria, 1968). Both cases have much to say about what memory actually is when we strip away our cultural intuitions.

S was capable of recalling sequences of numbers and words decades after a single exposure, and do so without any mistake. Whole matrices of numbers could be recited forward, backward, and by specific row or column. This capacity for manipulating memories far surpasses standard eidetic learning, which usually requires the recall to run in the same direction as it did when it was first acquired. S was aided considerably by his five-fold synaesthesia. Each sensory modality evoked the other four in some way, such that we get descriptions of things like this:

> When I hear the word green, a green flowerpot appears; with the word red I see a man in a red shirt coming toward me; as for blue, this means an image of someone waving a small blue flag from a window…Even numbers remind me of images. Take the number 1. This is a proud, well-built man; 2 is a high-spirited woman; 3 a gloomy person (why, I don't know); 6 a man with a swollen foot; 7 a man with a mustache; 8 a very stout woman—a sack within

> a sack. As for the number 87, what I see is a fat woman and a man twirling his mustache. (Luria, 1968, p.31)

Or this:

> For me 2, 4, 6, 5 are not just numbers. They have forms. 1 is a pointed number—which has nothing to do with the way it's written. It's because it's somehow firm and complete. 2 is flatter, rectangular, whitish in color, sometimes almost a gray. 3 is a pointed segment which rotates. 4 is also square and dull; it looks like 2 but has more substance to it, it's thicker. 5 is absolutely complete and takes the form of a cone or a tower—something substantial. 6, the first number after 5, has a whitish hue; 8 somehow has a naive quality, it's milky blue like lime…(*ibid,* p.25)

It seems that in each instance it is the sheer number of active regions which enables correct recall- if one is to fail, there are others to take up the slack.

Equally important is **S**'s description of how he would remember things. His early life consisted of attempting to create a whole scene of like for like visualisations of the events unfolding or being spoken about. This moving about internally was often exhausting, as each word or phrase needed some spatial element to it in order to fit within the storyline constructed.

S's extraordinary memory was entirely spatially oriented, but to such an extent it became hard to grasp which was real or not. So vivid and veridical was his recall that he could spend a day reminiscing about a previous day down to the last detail. Rather than a blessing, and one with which he intended to do great things with, **S**'s condition was more of a burden. Reading a story, or following a play was impossible without intense concentration. Previous experiences of the same or similar nature would invade his thoughts, and send his mind so far off from the actual words on the page he would not be able to even finish the sentence.

From the descriptions Luria provides, **S** appeared to be at the mercy of these episodes of mental time travel (perhaps here, mental space/time travel is more appropriate). In many ways, his internal processing dominates his thinking, and ignores the actual environment and he within it because of this overwhelming dominance. Total recall may bring with it enslavement to your own internal hypothesis mechanisms, such that you no longer require the world to necessarily be there and be sampled.

Such a concrete memory is one often seen in young children, where every little detail is seemingly attended to and noted down for potential future use. Abstraction, and the ability to link things without explicit ways of doing so are all properties of healthy adult brains, so it is easy to see **S**'s memory as being stuck in a developmentally early stage. A concrete memory did not mean **S** could not make mistakes. Especially early on in life, he was prone to miss

words from long sequences or omit details he was asked to recall. As Luria notes, a key difference with **S** was that these omissions were not as a result of forgetting the words or a failure in recall *per se*. Rather, the failure was in his *perceptual capacity*. Take for example this explanation of why he missed certain words from the same lengthy sequence of words:

> I put the image of the *pencil* near a fence...the one down the street, you know. But what happened was that the image fused with that of the fence and I walked right on past without noticing it. The same thing happened with the word *egg*. I had put it up against a white wall and it blended in with the background. How could I possibly spot a white egg up against a white wall? Now take the word *blimp*. That's something gray, so it blended in with the gray of the pavement...*Banner*, of course, means the Red Banner. But, you know, the building which houses the Moscow City Soviet of Workers' Deputies is also red, and since I'd put the banner close to one of the walls of the building I just walked on without seeing it...Then there's the word *putamen*. I don't know what this means, but it's such a dark word that I couldn't see it... and, besides, the street lamp was quite a distance away... (Luria, 1968, p.36)

If remembering everything is the result of a malfunctioning, as opposed to optimal, memory system, what of times when we fail to remember correctly? In short, the opposite. Misremembering is a sign of a memory system which is working optimally, one in which the when we look at it the right way, see that remembering is more of a fortuitous side-effect rather than the primary aim.

Most strikingly this can be measured in Alzheimer's and amnesiac patients. Participants were provided with a series of visual images, and then later asked to say if ones presented to them were on the original list, new but related, or new and unrelated (Koustaal *et al.,* 1999). When such items share the same thematic links, many people make mistakes. But amnesiacs shows a much reduced level of errors in tasks like this compared to healthy controls. Alzheimer's sufferers also show a slight but significant reduction in false recognition events (Budson *et al.* 2003). This can be interpreted as a reduction in misremembering and the accessibility to one's one store of semantically linked information.

A final study helps sum this all up. Participants were asked to recall autobiographical memories for specific events before being scanned and asked to imagine what would have happened in these events if the outcome was the opposite (e.g. positive to negative). The more likely this hypothetical could have been, the more the scans resembled activity during standard recall (De Brigard *et al.* 2013). What matters for the memory system is what is likely to happen or to have happened. When this goes wrong, we either know exactly what happened, sacrificing the ability to contemplate what might happen efficiently, or we are less susceptible to interference of memories, but only because we are impaired at learning anything new.

II.II Action-Oriented Thinking

So far, our memory system seems to be only working correctly when there is the capacity to misremember and to forget. There is also valid claim for navigation and **EM** sharing the same underlying neural base. I have presented evidence which solidifies this without making the link explicit. A recent study supplies some of this more substantial evidence, utilising place cell replay activity.

When recording from a population of hippocampal cells, Pfeiffer and Foster (2013) were able to decode the cell activity observed. They were able to use a Bayesian algorithm which was able to give the rough location of the rat given the cell activity within the environment. Importantly, it did so just from this cell activity. In the temporal window, the decoding also showed that in an open field environment place cells fired in manner which corresponded to paths through the environment. This occurred as expected whilst traversing the route, but more critically it did so before the rat approached that trajectory. The sequence was firing off in manner which may be interpreted as rehearsal, or expectation, even when not in the correct location for it to be realised immediately. Similarly, traversed sequences fire after in time-compressed, rapid and in reverse, and are carried by sharp-wave ripple episodes (Foster and Wilson, 2006). These are *replays.*

Replay in place cell networks provides a code for previously traversed trajectories and supplies an internal method of navigating an environment based upon egocentric prior experience. The place cells fire during periods of immobility (in a reversed, time compressed sequence) which when decoded matches to previously traversed routes within an environment, Foster & Wilson, 2006), or those yet to be travelled (Dragoi & Tonegawa, 2011). In addition, these replays have been shown to predict novel routes through a familiar environment (e.g. Gupta et al., 2010)[8].

All of this is possible thanks to the fine-grained spatio-temporal relations emergent from phase precession. In essence this network supplies us with where we have been, and when we were in a given place for a given trajectory (also the same for future trajectories). **EM** is richer than this, supplies our conscious thought with what we were doing this action for, the goals and objects within the traversed environment.

---

[8] Place cells firing sequences thus appear to co-vary with planning and analysis of actual movement of an animal. Yet it is still not confirmed if this activity is in any way casually tied to the animal's behaviour. Whilst unlikely not to be the case, we must be careful to not jump too fast to conclusions with work of this nature.

To link the two requires a task which requires specific encoding of what within trajectories to distinguish specific outcomes. Takahashi (2015) uses a subtask switching protocol- rats must navigate through a figure of eight maze and perform three separate tasks. Each task was for a fixed number of laps, but this number changed throughout a session. Therefore the rat could not necessarily predict the end of a set of trials. In addition, as this was all done within one session, any replay activity would need to be able to distinguish each subtask and its goals as independent entities if it was to be involved in memory encoding.

The subtasks- visually guided discrimination (**VD**), non-delayed spatial alternation (**NA**) and delayed spatial alternation (**DA**)- are sufficiently different so as to require some form of explicit encoding of the necessary actions required to receive the reward (in this case food). **VD** involved a light at a direction point indicating the correct direction. **NA** and **DA** had both lights on, and the rat had to alternate left and right for the food. **DA** imposed a further restriction with a barrier before the decision point for five seconds. This delay thus imposes on the place cell system a requirement to retain pertinent information on the task for longer.

These findings also support the previous replay work, highlighting that the replay that occurs is directly linked to the actual actions of the animal in the future. The state of the system of place cells decides prior to the choice point which is the predicted route to the goal, and thus the animal acts accordingly to this firing. This paper also picks apart the extent to which replays are specific to a given task type across the whole environment space. At all points prior to the decision point, the replay activity was for all potential task types previously experienced by the rat, not just the most likely. By maintaining all eventualities for the current environment of how one has to act, the system limits any chance of surprise. It additionally allows for seamless switching from one task-goal structure to another one.

As Takahashi himself suggests, these findings are indicative of replay possessing the same functional properties we assign to **EM**. If awake replay events can encode both the 'what-where-when' and 'what next', then given the other evidence we have accumulated there can be little doubt of **EM** being a capacity rooted in the navigational activity of organisms.

Replay is fast becoming the substrate for memory consolidation and retrieval within neuroscience. This form of memory planning and choice making however appears to have little to do with any complex processing in downstream regions, or deliberate consideration of facts represented internally. This recall is more to do with how were previous *actions* towards a goal successful, and how to go about replicating this. If there is any representing of

the world going on at this level, then it is fractured, overlapping, and heavily reliant on very brief time-compressed runs.

Time-compression adds another interesting note to the retention of information about events. Relatively distant events can be reactivated such that they become much closer to the goal they were responsible for achieving, enabling the decision as to whether an action is 'correct' or not. Often the causal linkages between our actions and the reward/consequence can take a long time to manifest. Thus this capacity to reduce the temporal element boosts our ability to learn, and is something successfully used in computer models of learning (e.g. Brown, Neath & Chater, 2007).

II.III Fast Sequencing Outside the Spatial Domain

These findings still require things to remain within the spatial domain. If **EM** from replay is to be the flexible and general memory we take it to be, then there has to be a manner in which this system transfers to the non-spatial domain. This could take a number of forms.

The first is that we are misguided, and **EM** in humans is very different from that in other organisms. That is, we have a system which encompasses all we have outlined, but does so in a manner independent of this older system. Redundancy is a key power of brain function, but a system with minimal co-opting of existing networks is wasteful and unlikely to hold up to scrutiny. A more plausible suggestion is this non-spatial experience is processed elsewhere and takes spatial information as a further loop within it. Thus we retain the navigational system as a bedrock, but this privileges the memory system as something distinct and 'higher' in a way which falls too easily into ascribing a more passive functionality to it.

The third option is that the same general system is used, i.e. replay-like sequences. Our brain is inherently a spatiotemporal patterned, complex bundle of cells that relies on both the spatial and temporal to generate the thought required to exist successfully within its environment. This direction is that the replay activity we see lends itself very easily to understanding navigation as the 'where' and 'when', even some 'what'. But this does not constrain it to just being used here. The general mechanism may underpin all forms of episodic memory we see, as an abstraction away from the necessity for actual movement and acquisition of object categories.

Evidence for this is slight. **MEG** studies in humans suggest a rapid generation of sequences akin to replay occurs in the hippocampus during a non-spatial navigation task (Kurth-Nelson *et al.,*

2016). This task involved movement between 6 picture 'states', where the aim was to move from one state to another in four moves to receive a potential monetary reward. The moves were not corresponding to any visible spatial layout, and the participants never saw or reported imagining such a setup. The **MEG** data matches with the rodent work, showing replay occurring in reverse and in a time-compressed manner. However, this **MEG** data may not have been located within the hippocampus directly, as the **MEG** signal arose mainly from visual processing areas. Whilst this signal may still be as a result of signalling between hippocampus and cortex (via **SWR**s), there is no robust evidence to support this.

Even if our capacity for non-spatial **EM** is not fully overlapping with that for spatial ones, the mechanism underpinning both is the same. Reversed, rapid, compressed, firing of previously active cell sequences seems to be key to learning, and to the capacity of an organism to know where it is and what it is doing in a conscious way.

II.IV Memory Traces and Forms of Representationalism

A memory trace as we have seen does not have to mean a mental representation, but is often thought of as being so. De Brigard (2014b) notes that in order for this relationship between trace and representation to hold there must obtain causal, retentional and similarity conditions. These are that the mental representation must be causally involved in event recollection, retain its original intentionality, and be structurally similar to the remembered event.

The causal theory of memory states that the initial event is causally linked to the recollection by the memory trace. This can be simply that the representation initially generated remains dormant and unchanged until recollection. This is a viable position within the theory, but empirically unsound.

A second, more intriguing interpretation is a reconstructivist one. That is, rather than storing the original representation- vehicle, content and all- there is instead some fragment which is sufficient to restore the original intentional content if activated during remembering. Even more specifically, one can be then content variantist- the intentional content of the initial experience does not have to be the same as the remembered content.

This violates the similarity condition, and also makes the claim that the memory trace thus does not have to match the recollection. All that is necessary is that it is causally involved in the representation that is recollected. Following this line even further, the three constraints

may not have to be properties of some neuron or cell population. Rather, they are dispositional- they come into being when the right conditions are met. A memory trace then has the disposition to generate an approximate representation with what is deemed to be the originally encoded content contained within.

Finally, we come to the difference between discrete and continuous representations. Discrete is that although we see our representations as continuous with one another, this is imposed by a secondary mechanism upon a set of individual mental representations. To hold the continuous position there is the rejection of this splitting up of memory into steps, and instead see it as a *process*. Unsurprisingly, we fast approaching[9] the view of Merleau-Ponty and Husserl, and thus the starting point for enactivist thinking. This is also the growing position in the memory sciences.

De Brigard goes on to illustrate the current theory behind memory research and the notions of consolidation, encoding and retrieval, much as we saw earlier. He concludes with the idea that based upon our empirical understandings, we cannot claim the existence of engrams as storing anything except dispositional properties:

> Finally, this scientific evidence also suggests that it is not accurate to talk in terms of a perception leaving a memory trace in the brain. Encoding does not seem to lead to any kind of brain modification that remains solely dedicated to coding for the stimulus that caused it. On the contrary, the neural networks engaged in the initial encoding get constantly redeployed in a variety of tasks and for a large number of purposes, which likely cause them to undergo further transformations. As such, it may be best to think of a memory trace, not as a neural network constantly carrying a particular memorial content from encoding to retrieval, but rather as instantiating the dispositional property to reinstate, as closely as possible, the complex hippocampal-neo-cortical pattern of neural activation the brain was in during encoding, at the time of retrieval. (De Brigard, 2014b)

I am inclined to agree. Synaptic changes in protein expression which occur with the generation of a memory are not specific to that one memory. The cell networks we call engrams or traces are used in other assemblies for other memories or cognitive functions. Dispositional properties are efficient as they allow for reuse of neurons, whilst still providing a reliable ability to instantiate neural activation patterns.

II.V Schemas and Association

So far the focus has lied on **EM** in a basic form. Another important aspect of our declarative memory is our **SM**. If we take the position that **SM** is a specialised abstracted form of **EM**, then our account of memory will have to incorporate this idea at least tentatively into the framework.

---

[9] Although, admittedly, not quite *reaching*.

SM is deliberately quite removed from plastic and highly active brain regions. Whatever we call the trace or information, it has been separated from our episodic, spatially centred hippocampal region in order to maintain some stability. The question we have to solve is just why this is, and does this commit our account to a representative explanation at this point, or can we still make use of dispositional properties.

One claim may be made, that **SM** is a distinctly human trait, and thus representational in virtue of language or our capacity to acknowledge it as having content. Thus we just detach this part of memory from the rest, and concede this as a complex form of cognition that transcends our enactive approach. This would be reasonable, but a little defeatist. Evidence for particularly complex **SM** is evident in rodents (Tse *et al.,* 2007, 2011) and primates (Osvath & Martin-Ordas, 2014). The argument appears to take Tulving's (2002) suggestion that declarative memory is only recently emerged out of human evolution at face value, ignoring that animal work has been the driver of our theoretical considerations in memory for the past few decades.

Suppose you move to a new city for work, and are invited on your first day to eat at a restaurant a mile or so away. When you set out to find this place, you will most likely resort to using maps, GPS, landmarks and asking people for directions. If you do none of these, your chances of finding the restaurant are slim at best, even if you knew the street name. Yet over time you become more confident in your ability to find the place from a multitude of different starting points. After a while you go to other places, which again take some time, but you learn their location and eventually no longer need to really think on how to get there.

Now, imagine a new restaurant opens a year or so later. Now it is unlikely you end up getting lost or having to ask for directions. Your knowledge of the city, the navigational mastery you have will enable you to find this new restaurant without any major trouble. You may even use the other restaurants as reference points on your way there. Yet you no longer remember all the previous routes you took in an episodic fashion (unless they were interesting in some way). Instead, you have a *sense* of where to go, built up over time from repeated exposure to navigating through the city.

This bank of knowledge is a *schema*. It does not require rich, explicit memories of exactly how you have previously got somewhere, but the general *gist*. A gist is in most respects the same as a hypothesis, it predicts the correct response (or more often the correct meaning) given previous experience. The associations made in navigating around the city provide you

the capacity to know how to get to novel locations by predicting the likely correct route to take.

This was empirically shown by a number of studies using an open maze environment with rats (Tse *et al.,* 2007, 2011). The animals are placed into one of four boxes at the edges of the 'event arena' and provided a food pellet. The flavour of this pellet corresponds to one invariant location within the arena out of six. After finishing the food, the rat is allowed to run into the arena and dig to find the remaining food. Thus the rat develops a paired association of flavour to a given place, and must acquire a reliable way of remembering this for subsequent testing. In a day the animal will be tested on all six possible pairings. This necessitates a mastery of how to navigate within the environment using only sight (visual cues are the only ones available to the animal to complete the task).

This complex task takes over a month of continuous training to acquire with few mistakes made on a given trial. Yet once this is complete the animal is able to find the food rapidly, without much deliberation or exploration. Then two locations are removed, and two new ones added with new flavour pairings. Rather than a month, these new mappings are made within one day. The hippocampus has rapidly extracted that this change to the task relies on the same structural relationship as previously. Thus there is no need to retain individual experiences of this new exposure. What is necessary is to situate each location within the environment and in relation to the others still present.

Remembering exactly what happened here is completely superfluous to success here. After this single day of exposure to the new pairings, the hippocampus can be lesioned bilaterally. After a two week recovery period, the rat will still perform as well on the new pairings as the old. A schema has been constructed over the course of the initial training which allows for rapid gist extraction and subsequently no need for experiential memory to be maintained. New pairings post-surgery are unable to be learnt however, reflecting the idea that at least some of the encoding of a memory trace is necessary for this extraction process.

This memory trace encoding can just be the dispositional properties we encountered earlier. The hippocampus is well placed with **EM** to extract dispositional properties and thus reduce things down to their gist. The further systems consolidation to neocortical network states is a further compression of these properties made possible through similarities to other already extant **SM** traces. We will return to this in **Section III.III**. But first, we should look further at how this idea of schema came into being, and how it relates to human memory.

II.VI Bartlett and Remembering

This idea of a 'schema' stems from pioneering work carried out a century ago by Frederic Bartlett. He was the first to define the term as we have used it, capturing the elements of reconstruction, gist, and action which lie at the heart of this reimagining of memory.

Bartlett, in his 1932 book *Remembering,* provides a neat example of how memory of a specific passage becomes over time less rich in detail, depending on the number of times the material is recalled. Individuals were asked to read an old obscure American folk-tale, *The War of the Ghosts.* After reading through twice, each person was asked to recall the story at different intervals, without being told why the experiment was being conducted[10].

What was found was that the story was often quickly reduced to a 'journalistic' style- awkward place names were omitted, conversations stripped of most of their content and many misunderstandings of the story made prominent. Over longer timespans of recall these mistakes and the person's own cultural background begin to override the actual tale. Rationalisations of the ambiguities in the story are clearly there to make sense of the story *for the individual.* Each person has a different take on the central message, what details are important, and how to structure the story. The subjective aspect of memory, of how our experiences always influence us even when attempting to be objective, is clear here.

One person with a specifically visual method of recall showed a clear grasp of the gist of the story for a long time. Over two years after the read-through, all that remained however was the barest outline, one where there were remembrances of personal elaborations over the original story. The outline bears the major points, of not correctly at least in close relation. As I have argued, this is what memory functions as, a way of extracting the gist to better suit hypothesis testing for future events. To illustrate this further, Bartlett also asked for a recall of the story in another two people at even longer intervals. The first, at six and a half years after, showed only semantically related elements of the story. The whole was constructed based upon what he *imagined* was correct- he believed it to be fairly accurate, but aside from the key plot element the rest was his own. The second, at ten years after, could remember two place names, and a vague visual image, but nothing more. These names, according to the person, had been prompted by seeing Bartlett in the street two years after the test- the presence in the environment of something related to the task configured her memory system

---

[10] Whilst an old and not rigorously controlled experiment, Bartlett's evidence is rich in the kind of experiential detail lacking in much of modern day research. In embracing phenomenological approaches to cognition I feel it is good to return to these old studies in order to see what can be captured by similar types of approaches as are suggested by people such as Varela. For a similarly eye-opening take on how education alters our cognitive faculty, seek out Luria (1976)

such that some information was generated. This is more readily noticeable in the 'Doorway Effect' (Radvanksy & Copeland, 2006; Radvansky, Krawietz & Tamplin, 2011). Walking into a room you have just been in often is enough for you to recall something that had been a few moments ago completely irretrievable. The context of our initial generation of an episodic episode is quite important in assisting in the learning process.

We often immediately condense and simplify our experiences, or give one element prominence at the expense of the rest. Our living *now*, in the present, is the only time we possess the experience. Past this, we are dealing with the indicators of this experience, what information is left in the system state configuration changes made at the time. This however makes remembrance nothing to do with the recall of content as it was, only the remembering of what the content may have been. The form of the content is maintained, the structure is evident, but the veracity of the generated content upon recall matches because of semantic knowledge, not any store of the original content as it was. Most changes in what we recall happen in the first few remembering episodes, emphasising this lack of true storage.

Bartlett was clear in his assessment that this work showed memory was reconstructive and active.

> Every story presented had to be connected, certainly as a whole, and, if possible, as regards its details also, with something else. This is, of course, the factor which I have already called 'effort after meaning' coming again into play. It could be said that there is a constant effort to get the maximum possible of meaning into the material presented. So long as maximum of meaning is understood to imply an effort to find that connexion which puts a subject most at his ease in reference to a given story, the statement is true. The meaning, in this sense, however, may be of a very tenuous and undetermined nature, and apparently may even be mainly negative (Bartlett, 1932, p.84)

> [...] the process is emphatically not merely a question of relating the newly presented material to old acquirements of knowledge. Primarily, it depends upon the active bias, or special reaction tendencies, that are awakened in the observer by the new material, and it is these tendencies which then set the new into relation to the old. To speak as if what is accepted and given a place in mental life is always simply a question of what fits into already formed apperception systems is to miss the obvious point that the process of fitting is an active process, depending directly upon the pre-formed tendencies and bias which the subject brings to his task. The second point is that this process of rationalisation is only partially —it might be said only lazily—an intellectual process. No doubt the attempt, however little defined, to seek out the connexions of things is always to some degree intellectual. But here the effort stops when it produces an attitude best described as 'the attitude in which no further questions are asked'. The end state is primarily affective. Once reached, and it is generally reached very quickly, it recurs very readily, and it is this, more than anything else, which accounts for the persistent sameness of repeated reproduction. (*ibid*, p.85)

Our memory strives to rationalise our experiences into something coherent and viable given our previous experience. Once a story is in place, the closeness to other system states is what stabilises it, how we avoid losing the gist of the experience even if some of the context alters or is lost. There is no fitting of this new information into a rigid immutable framework of past

knowledge however. The past influences the perception and actions carried out in the now, which thus alters what we take as important, and thus how we will potentially act in a similar situation again. Much like enactive accounts of perception, our enactive account of memory finds no need to store anything exactly as is (or was).

A few paragraphs of *Remembering* succinctly highlight that Bartlett was a dynamicist in his thinking on memory long before it was taken up seriously elsewhere:

> …though we may still talk of traces, there is no reason in the world for regarding these as made complete, stored up somewhere, and then re-excited at some much later moment. The traces that our evidence allows us to speak of are interest-determined, interest-carried traces. They live with our interests and with them they change (Bartlett, 1932, p.211–2).

> I have never regarded memory as a faculty, as a reaction narrowed and ringed round, containing all its peculiarities and all their explanations within itself. I have regarded it rather as one achievement in the line of the ceaseless struggle to master and enjoy a world full of variety and rapid change. Memory, and all the life of images and words which goes with it, is one with the age-old acquisition of the distance senses, and with that development of constructive imagination and constructive thought wherein at length we find the most complete release from the narrowness of presented time and place (*ibid*, p.314)

Our memory is special only in that it allows us to enact the environment outwith the current spatiotemporal location. It is a function that predicts and strengthens hypotheses without being one of stored content about those past events. If it is a store, it is one of patterns of activity, which may as an emergent property give rise to recall of the events which actually happened. We 'master' the world by being able to anticipate quite well all possible outcomes, and make reasonable assumptions at what did occur. We can do so because memory is not some isolated system that only receives input and regurgitates it when required. Declarative memory systems actively shape any input, and modify it based upon the whole state of the organism, such that future access to this system state generates actions likely to be accurate and successful.

II.VII Flashbulb Memories

So far I have painted memory to be an unreliable narrator, supplying what it thinks was present depending on the current biases of the cognitive agent, and the dispositional states it possesses. Yet we are all acutely aware of some occasions where memories are recalled in vivid detail. Every action and object is present, even smell and sound appear to be retained and brought forth again *exactly* as was the first time round. Nearly always these memories are so strong because of the emotional impact they have, and are most commonly associated with shocking events.

These are referred commonly in the literature as 'flashbulb memories' (Brown & Kulik, 1977), a snapshot of a moment in time where everything was illuminated so brightly that every aspect is captured. This is attributed to the importance of such events- it may be useful to recall everything as there may be key details that allow you to survive or to make sense of what just occurred. To accommodate for this, all sensory input over a period of time during and after the event is somehow retained permanently, bypassing normal encoding and decay mechanisms. Yet if this is the case, then enactive memory runs into problems.

If memories are capable of being actually stored in an immutable fashion, then the enaction of memories and the world is nothing more than a cost saving exercise. Yet the enactivism I am proposing does not allow for such permanent traces within a healthy functioning system. Thankfully, this tension can be resolved through investigating the nature of these flashbulb memories a little deeper.

We ascribe a sense of certainty to the recollection of flashbulb memories as being the exact replica or re-experiencing of the event, as opposed to when we recall our breakfast from the previous week. Flashbulb memories are given a high accuracy rating. We never stop and doubt our experience, nor are we very likely to be manipulated into altering it. This memory however may still be incorrect. The accuracy we rate this memory with masks our ability to assess the validity of its contents. In short, these memories are no different to our everyday ones in how they function. What has changed is our acceptance of them when we recall them as never changing with time.

Our memory system is thus enacting these 'strong' memories, providing so much perceptual detail we are convinced of our own self-generated experiences as being veridical. This must pertain to some form of accuracy conditions being applied to the experience, as opposed to truth conditions. All memories are 'true' in the sense that they are correct for the individual. Only an outside observer could define the truth of any single memory, and even then be subject to their own internal accuracy conditionals.

Immediately after the 9/11 attacks, students at Duke University were asked to record the experience they had on first hearing the news, and then tested on how well they recalled these details either 1, 6 or 32 weeks later (Talarico & Rubin, 2003). There was no difference in the rate of forgetting for these memories, nor in how consistent each recall was compared to the original. It appears that the level of tiny details supposedly remaining within a flashbulb memory is sufficient to convince us of the accuracy of the memory. This ties in with De Brigard's (2014a) idea of episodic hypothetical thinking- we think it likely to be the case so we

take it to be correct because of its consistency with our held schemas. Once again, these memories are not special compared to everyday memories, only are judgements on them are different.

III. Enactive Memory

Thus far, I have sketched out a description of memory which goes against the standard default we fall back to in everyday usage, emphasising the evolutionary heritage of **EM** within navigation. I have yet to lay out plainly how **AE** assists in this endeavour however. Much of this work makes claims of interaction and dynamicism; why add another paradigm into the mix at all?

Memory did not arise fully formed from the depths of human- consciousness. It is hard to even conceive of it as a single entity, yet we persist with the idea we can study memory as a system which is shut off from the rest of the body and environment once it has taken in the information it corresponds to. We cannot rule out the position that in some cases, thanks to our language faculty, this is actually possible. What we can argue is that these instances are ones of malfunctioning, faulty memory systems, and are in fact detrimental to the cognitive agent. Forgetting is a vital part of learning and memory, without which it is hard to make linkages which are not concretely experienced.

The initial experience is inherently an interaction between body and environment. What emerges from this is a tagging of a particular state of the system that indicates the experience. Thus a memory indicates an interaction has occurred or is occurring, it emerges from this interaction. The capacity for memory is not just there, and then filled in when necessary by sensory input. Rather, it is generated by the interaction with the world. This is why I feel that an enactive memory (**EnM**) is necessary. Only by combining dynamical principles with a focus upon the importance of the organism will memory research be able to fully appreciate how our **EM** and **SM** function.

III.I Cognitive Maps, Enacted

To bring our enactive memory account into being by building from the ground up, we need to tackle the **CM** head on. As a highly representative and immensely useful model of neural activity, we are best if we maintain it as a viable way of describing what goes on when we navigate through the world around us. Yet we are committed to embedding this into action and interaction much more keenly than is currently presumed.

In order to achieve a reconciliation, let us make use of an example. Animals navigate by using cues as ways of carving up the environment such that areas can be recognised as being informationally rich or not. If we take a large distant cue, such as a castle or tower, we can use our position relative to this cue to calculate where we are and our own position. Now in our traditional **CM** picture this involves the encoding of our own motion and heading and marrying this with information of surrounding cues and their relative positions to us and to other cues. To do so requires communication between brain areas and the generation of an internal topological map which can be referred to during planning. This information is then utilised in our activity to provide us with accurate potential trajectories to reach our intended goal.

In an enacted world much stays the same. Except now the focus shifts away from an internal 'map', and focusses on what is happening in real-time as we move. The firing of our navigational subsystems is all to maintain an understanding of where the organism is *in relation to everything else.* By acting and receiving feedback the organism modifies its behaviour accordingly. We as humans ascribe conscious categories and names for objects, define things specifically as landmarks or cues. How much this extends down to other species is unclear, with semanticisation like this a level above our episodic focus. It points however to an important observation that when we navigate we do not depend on being able to identify or categorise the cues we use to navigate.

For our purposes of acting in the world, it is sufficient to know that a region of our visual field remains fixed even when we move substantial amounts, and compare this to other areas that do seem to shift in our perceptual domain. Tagging these with definitions and names certainly helps in our ability to describe what's going on, and may even aid in the activity itself. But it is not necessary to know anything about what we are using to navigate except its relation to ourselves and regions around it.

Enactive thought then is the directedness towards elements of the environment which in turn direct us to other parts. There is a rudimentary sense of assigning some knowledge of certain aspects of the environment, as we must know how to act given a certain configuration. But this is what our memory is geared towards primarily. In a general sense, with the procedural memory of how to perform explicit actions, and in the specific sense we have discussed of trajectories and goal achieving.

These aspects are, crucially, the patterns of activity that result in specific types in interaction with the environment. These patterns are dependent upon the environment itself just as much

as the underlying neurons. Nor is a pattern strictly fixed to a given cell or set of cells, as we have seen the brain utilises neurons for multiple roles which often overlap. An internal model with fixed neural patterns would work just fine, but require a brain much larger than our skulls are capable of holding.

Perhaps **SM** is evidence for some fixing of patterns into certain neuronal assemblies. But even here the conditions must be right for these neurons to regenerate the correct activity profile that brings about the relevant information. So too, is the likelihood these cells are reused elsewhere. If there is any retention of a representational nature, it is slight.

III.II Autopoiesis and Memory

Autopoiesis is a theory of organisation, not of function. So when we speak of memory and its relation to autopoiesis, it is how such an organisation of an organism allows for the functionality that we place on memory.

When speaking of cognition within autopoietic bounds, it is best to give a clear idea of where the theory lies on how we experience the world. Varela, Thompson and Rosch give three examples of how may think of cognition. The first presumes the world has properties already without perception from organisms such as ourselves. Our cognitive system attempts to access these properties and use them by representational ways of thinking. This is a cognitivist claim. The next is idealistic- the cognitive system projects a world of its own making, with all rules, laws, properties etc. coming from within the system itself. We loosely here have an objective/subjective dichotomy at play.

The third example, the one which we shall take as that best fitted to **AE**, lies down the middle of the previous two. It holds that, unlike the objective view, categories are an experiential construct dependent on the history of the organism and the socio-cultural environment it currently resides in. Unlike the subjectivist view, this does not make these categories wholly internally generated; they are part of the shared cultural and organismal world.

There are three minimum conditions put forward to be an enactive agent:

1. **Individuality** An agent is responsible for producing as well as maintaining its organisation.

2. **Interactional Asymmetry** Based on point 1, the agent must be the one that controls any parameter changes to the coupling with the environment. It must be the *main active player* in the interactions.

3. **Normativity** The agent creates norms for the interactions dependent on the activity that is carried out.

It is these three conditions which set the **AE** position apart from others such as **DST**. Thompson (2011) makes clearer the nuances between **AE** and **DST** in a reply to Oyama. In it, he describes that in **AE**, there is indeed the idea that an organism enacts the environment, which modulates the organism as this happens. Both are necessary for sense making, and neither is sufficient alone. Where **AE** differs from **DST** however is that the 'inside' of the boundary (that of the organism) is ontologically prior to the 'inside-outside' dichotomy. This is what creates the asymmetry Thompson sees as a necessary condition for agency.[11] In qualifying this point, Thompson refers to work by Moreno and Barandiaran (2004). They attest the inside is the creator of the outside, as the organisation of this 'inner' is what controls the interactions and relations. However, both environment and organism are necessary, and neither are sufficient by themselves for sense-making (Welton, 2011).

An enactive agent can thus alter the environment to suit its own goals, or make use of the current state preferentially[12]. Without interactional asymmetry, our sense-making would not be able to relate to the intentionality of the organism (Thompson, 2011). Intentionality and normativity are key features of **AE**, which distinguish it from **DST** or autopoiesis alone. These are bound up into autonomy of the organism. This entails self-determination, that the organism generates itself. It also, in the **AE** case, refers to the operational closure of the system. This is simply that the systems processes are circular- there is no way to define a start or end which is meaningful.

Normativity arises out of the individuality (autonomy) of the enactive agent. For Barandiaran (2016), autonomy provides a way of naturalising normativity. He refers back to *The Embodied Mind,* and how much of the work there emphasised the notion of autonomy (as operational closure). Varela's use of autonomy is what distinguishes it from earlier forms of autopoiesis, and the notion of normativity included therein. Barandiarian claims that:

---

[11] Note that this does not mean the boundaries are created externally to the dynamical relationship, they are fully there because of the dynamics.

[12] Thompson (2011) here gives the example of a bird gliding. Bird flight makes full use of thermal updrafts and air currents to stay aloft. Flapping wings is only necessary when these all fail or are insufficient. The bird surfs these environmental features and use them to their own ends.

> Autonomy anchors normativity in the large-scale plastic correlations and homeostatic inter-dependencies of sensorimotor coherency, instead of relying on linguistic, evolutionary/adaptationist or representationalist principles. (Barandiarian, 2016, p.19)

Norms may be shaped by social practices, but are based upon the dynamics (and more specifically neurodynamics). The coherence levels of dynamics inside the organism create norms through emergence, still dependent on coordination with the environment (Barandiarian, 2016). Our little bacterium still swimming within the sucrose gradient has a capacity to behave as it does through its sensorimotor coupling. These quick changes in response to the environment slowly move the dynamics of the system into an organisation moulded by these patterns, reinforcing some areas, weakening others and reshaping the meaning of things in the environment. Norms or habits of the system arise from the need of the system to behave in order to maintain the viability of behaving.

The intentionality within **AE** is best affirmed as directedness, a way of engaging with the world that is co-constructed with the enactive agent. Because of this co-construction, the state of the organism defines the potential interactions with the environment, and how these interactions will go. Elsewhere, Thompson defines cognition as enacting a world through a history of structural coupling. This history only needs to be viable, it does not require optimality. In other words, any action as valid as long as it does not threaten the integrity of the autopoietic system. The organism in its individuality creates the importance of certain parts of the environment because it is limited in how it can direct its actions.

In outlining **AE** initially, I made reference to the idea of mechanistic principles. Whilst the pure form of autopoiesis claims to be mechanistic, Varela's own take in **AE** appears to not do so. There is a rejection of making organisms machines in this way, and an emphasis on the organism's interactions which appears to make mechanistic approaches untenable within it. In *The Embodied Mind* however, there is clearly laid out an example of how to understand colour which appeals to dynamical explanations which are 'extensive'. The mechanism is not just the organism, but the environment, social and cultural spheres it currently inhabits also. There are commonalities shared with all organisms, but our explanations still include individual-dependent relations with the world (Rucinska, 2015).

Carlos Zednik (2011) seems to think this is possible, by using dynamical explanations for the distributed and multi-layered mechanisms necessary to adopt in cognition. He suggests they are "uniquely able to describe mechanisms whose components are engaged in complex relationships of continuous reciprocal causation" (p.239). Varela suggests our understanding of colour in enactivism is doing something of this sort:

> …the overall concern of the enactive approach to perception is not to determine how some perceiver-independent world is to be recovered; it is, rather, to determine the common principles or lawful linkages between sensory and motor systems that explain how action can be perceptually guided in a perceiver-dependent world (Varela, Thompson & Rosch, 1991, p. 173)

Colour is part of a history of structural coupling which involves cultural as well as biological components. Rucinska (2015) thinks that this description appears to conform to the idea of an extensive mechanism using dynamical explanation being possible within **AE**. I feel this is a way for **AE** to find better fits with empirical work in neuroscience, even if it does leave it open to the threat of someone suggesting these mechanisms are representational.

Taking all of the above, we now must marry **AE** with our sketch of memory and see how well the relationship holds together. Imagine two people view the same series of events unfolding in a room, make the same movements and say the same things. Each person will take away from this a different set of memories of what happened. This is to say, memories are approximate and dependent on the system state of the agent. A memory is not an isolated unit, it is set into the world around it by virtue of being an enaction of a potentially previously experienced state. In addition to individual state differences, cultural differences will alter the perception of the scene and thus any memory which may be later reconstructed.

When we remember something, it is not because the memory was already there somehow in the world or in our heads, we create it by virtue of being an active autopoietic system within an environment we have helped create. In addition, we are influenced by our own creating of this environment as well as by other agents which have done so, such that we enact a world similar to that of those around us.

Our memory, therefore, is not something inner, a directing of acts *inwards.* The retention of previous interactions comes about as a result of the operational closure- the start and end points we ascribe sometimes (e.g perception, action, encoding) make sense only to us as observers, not to the system. Thusly, our recall of past events should not be seen as justifying the claim that declarative memory is only about remembering the past. It may not be primarily about remembering what happened as a continuation of conscious self-hood, but instead a method of extracting the patterns that were encountered (or were *viable*) in order to further future activity. The creation of a self with the ability to consciously reflect on this was a secondary emergent property.

Accepting a wider definition of mechanism also allows for us to include empirical work from within neuroscience which is committed to a neurorep and (sort-of) mechanistic view without contradicting ourselves with rejection of mindreps and accepting a traditional mechanistic stance[13]. So what of this notion of dynamics? How do cells with firing that matches onto features of the environment act in way explainable through dynamical means?

III.III Dynamics of Oscillatory Activity

A problem faced by any attempts to investigate cognition neuroscientifically is to close the gap between the neurobiology and the phenomenological features of our thoughts. Varela proposed a field of neurophenomenology would be best poised to take up the evidence provided by neuroscience and interpret it with regards to the work of Husserl, Merleau-Ponty, and that of autopoietic theorists (Rudrauf *et al.*, 2003; Varela, 1996, 1999). The essence of what is attempted within this field is to 'bridge' this explanatory gap using neurodynamics combined with subjective experience (see Thompson, Lutz & Cosmelli, 2005 for an excellent overview). Whilst not taken up to the letter (for example I do not find a convincing argument for how best to analyse first-person phenomenological data), it has been in spirit taken to heart here.

What is made clear in Varela's reasoning is that our cognitive processes should be thought of as operating through trajectories in a 'phase space'. This is a multidimensional entity in which all possible states of the system are laid out. Thus given knowledge of the current state of the system (time t), it is possible to reliably predict where it will end up at $t+1$, and where it has been ($t-1$). In some systems, any starting state always results in the same final state- these regions are known as attractors, as when drawn out the trajectories of the system seem to be pulled towards them. There is a whole host of complexities to how these attractors may function past a simple point attractor. The reader is directed elsewhere for explanations of these (see Izhikevich, 2007 for a good primer). For our purposes, it is sufficient to understand that the initial conditions of a system lay before it a series of paths which are determined as long as no external input to the system occurs.

Earlier I outlined the population of cells which in some way co-vary their firing with specific properties of the organism with respect to its environment. Just *how* these cells acquire their complex firing properties however is critical to our emphasis on enacting memory, and to understanding if they possess dispositional properties. A purely representational form would be that they acquire the properties as a result of the presence or absence of features, and that

---

[13] I admit this notion of 'wide mechanisms' needs further work, but this lies outside the scope of this paper.

these are acquirable without the need for the activity of the organism as a whole. It would be sufficient to have inputs from sensory modalities which corresponded to the set of features the cell responded to.

Most of these cell types are modulated by theta oscillations. Originating from the septal, hippocampal and thalamic regions, theta rhythm tends to lie between 4 - 12 Hz. Theta-generating cells have one specific property that links them to the body as a whole- their firing correlates with the movement of the organism. Without movement, there is little to no theta rhythm. It has been proposed that theta cells are the primary modulator of navigational and memory cells because they specifically modulate themselves so as to encode speed and direction (Welday *et al.* 2011). These are the ones which drive the theta rhythm within the hippocampus and **EC**, with major populations in the medial septum and hippocampus.

Theta cell bursting fits tuning curves which are then capable of entraining the place, grid etc. cell populations to their respective roles. The firing activity of these populations is thus extracted from a specific matrix of theta cells. Each cell will fire only when theta cells from this matrix are synchronous, which corresponds to a correlation with a certain speed and direction. These findings suggest rather than being a set of different cells which encode for information relevant to one another, all of the navigational cell types found take their firing profiles from the same source. This goes against some of the traditional stances (but see Langston *et al.,* 2010), yet may not be the whole picture.

Arranged into ring attractor networks, these theta cells act as velocity controlled oscillators (**VCO**s). The grid cells, place cells, etc. take input from different sizes of theta cell populations, and wait to detect coherence (that is all the inputs fire at the same time). Upon detecting this, they then fire. A cell that only receives inputs from a few theta cells will fire multiple times in an environment, much like a grid cell does. Taking inputs from a large number of **VCO**s however will only likely result in one firing location in the environment, which resembles that of a place cell.

These **VCO** rings only have one cell firing at any one time. The cell which fires moves around the ring dependent on the theta frequency. Each ring only has theta cells which have the same direction and speed preferences, whilst between rings these preferences differ. Such a system is at once powerful and non-specific. Firing profiles are only dependent on the theta cells that are sampled from.

This is unlikely to be the only driver of this system however. Theta cell input is not sufficient for the normal operating of these cells. The connections between navigational cell types is vital in maintaining their stability (Van Cauter, Poucet & Save, 2008), and may not be fully explainable by **VCO**s alone. Something else appears to be necessary for the system that is independent of the **VCO**s. Medial septal inhibition does not affect place fields drastically. The firing *pattern* is altered immediately, but the fields remain stable and do not become theta-modulated again upon reactivation of the medial septum (Brandon et al., 2014).

The theta rhythms themselves may not be required in all cases for navigation (Orchard, 2015). Bats lack any theta rhythm at all (Yartsev, Witter & Ulanovsky, 2011), and humans have it much less commonly compared to mice and rats (Jacobs, 2014). Supporting this is a very recent paper which has uncovered a population of neurons in the CA2 region of the hippocampus that fire when an animal is immobile. In doing so they appear to maintain the stability of the place cells even in the absence of theta rhythm (and are perhaps linked to SWR events) (Kay *et al.,* 2016). Theta cells it seems are necessary for egocentric navigation (path integration and grid cells, but allocentric does not require them to exist.

Beyond navigation, into memory, we can see how such attractor networks may be utilised to recall information. A similar set of initial conditions as experienced previously is likely to result in the same end point, such that the same experience occurs that did so before. Memory is then the simple returning of the system to previously held states within the phase space. But how does this occur when we recall memories long held, or for locations/situations that no longer exist? We cannot enter a similar state by being in a similar environment, as it may be impossible to do so. Somehow we must possess a capacity for returning to these states in a dynamic manner which is not reliant on this jogging of the system.

AE claims that as enactive agents we are capable of controlling our environmental access and ride the dynamics of it to benefit the organismal side of the enacted world. Within the nervous system, which is itself autopoietic, there then may be a way for the memory system to ride the trajectories in phase space to previous experiences. The first thing to occur is that of associativity. Our semantic learning capabilities often rely on drawing linkages to other information that we hold. We also possess models of this learning which place it within similar multidimensional dynamic spaces with distances between information an indicator of similarity. We are rarely asked to recall information in a contextual vacuum. This information our surroundings or just-experienced events provide may be enough to allow our memory to ride through phase space to regenerate old states.

This has been suggested more robustly as a form of chaotic itinerancy. Here attractor states are memories, but transitions are dubbed chaotic. Input to the system which may be new causes the activity to 'jump' from state to state of 'known' ones, as if searching for a match. If known, the system converges quickly to the correct attractor. These attractors are better named as 'quasi-attractors'. They are regions of convergence within a phase space, but only weakly, such that any orbits are of such long periods they are unlikely to be observed experimentally. This has been said to resemble "deterministic chaos" (Infeld & Galkowski, 1997, p.56-7); the regions of (weak) attraction lead to areas of widely divergent and chaotic activity (Tsuda, 2015). Thus the system appears to move in a random fashion, but is fully obeying a set of nonlinear dynamical equations.

Another way of thinking of this is when there is a group of these quasi-attractors, akin to a set of memories. The nature of these attractor states leads the area to be both attractive and repellent simultaneously for the system (Ishikawa *et al.*, 2007). When the system finds itself near an attractor ruin, it is attracted to this dynamical configuration before being repelled away back into areas of instability. These attractor 'ruins' are destroyed as soon as they are exited; the system maintains a constant searching through phase space (Harter & Kozma, 2005). This is much like our conceptualisation of spatial memory. The system is constantly moving through attractor ruins into other states dependent upon how the environment sets constraints through recurrent interaction with the organism. One ruin sets the system to search for the next one, which resides as a dispositional property within cell populations of the hippocampus and neocortex. 'Drifting' through these states in the correct context allows for our ability to recall the echo of a previous experience.

By adopting chaotic itinerancy, we take a dynamic position which is always unstable, always reaching and grabbing for the next attractor ruin. These ruins are transient weak convergences which are destroyed no sooner than they have been traversed. There are parallels with enactivist account of perception here, and so too with our memories. We know how to get back to similar areas of the phase space because the system is deterministic, and we only require that the system is disposed to returning to a similar state.

III.III.i Information foraging

Johnson *et al.* (2012) present a theoretical account of the acquisition of a schema that ties in with our **EnM** account and with the dynamics outlined above. Initial exposure to an environment generates a sampling plan which prefers areas which are highly informative (i.e. densely filled with information that may aid action in the environment). This information foraging allows for improved searching and goal realisation within the given location- i.e.

much like areas of meaning in **AE**. Information-rich areas are calculated with the Kullback-Leibler divergence. This states that given the set of hypotheses of locations that are interesting, there is a value of which one will provide the most information. In the paper, these are 'observation functions': the probability that at a location L an observation is made, assuming a desired goal is present within the environment. Goal locations therefore can have prior probabilities attached if there is a bank of prior experiences that found the goal in that location.

This kind of searching however does not require any representational element to be existing in the navigational system. **AE** in this manner has no need to utilize any complex coding within **CM** cells other than where the animal is at any present time. All the cognitive legwork is done elsewhere upon these prior probability distributions. Johnson refers to this as undirected foraging- there is no utilisation of the prior probabilities (if we assume these to be memories) in the actual foraging behaviour. Directed foraging involves making decisions about which location(s) to sample, testing hypotheses in real-time via hippocampal sweeps of potential trajectories. Behaviourally, this is seen as looking back and forth between them, known as vicarious trial and error (**VTE**). Sweeps occur simultaneously with the **VTE** behaviour, in effect indicating the path being looked at as having some value at succeeding in the goal at hand.

III.IV Distributing Memories

If we wish to move away from our memory as a store of experiences, then some of the load must be distributed out of the neural system into the wider world. We have already seen how enactivism can bring about this with regards to the environment. Yet we have ignored thus far the extent to which other organisms, tools, and the dynamics of society may assist in this. This is to delve into territory most commonly attributed to extended mind theories such as that first postulated by Chalmers and Clark in 1998. We committed earlier to the idea that our memory traces are dispositional states which act to generate patterns of activity, as opposed to discrete memories. These patterns supply the neural basis of recall, but not necessarily enough information to accurately replay the same content of the initial experience.

Imagine a car crash happens on a busy road filled with shoppers. When multiple people are within the same place and experience the same event, then we have a greater chance of accurately being able to recover what actually happened. But if we were to take just one or two accounts and compare them, we would likely find many inconsistencies and outright contradictions. The information is only ever approximate because of the lack of any commitment to truth within our memory system as we have sketched it. Summing over

multiple accounts allows for removal of 'false' information and retains only that which is most likely to have occurred. We cannot of course say this memory constructed from many is objectively truer than a single person's- the middle way also reminds us that the subjective experience of a person will always play a part in the recall. The *accuracy* of the report however is strengthened.

Societies provide an external format where patterns of activity can be maintained through inter-organism dynamics. A crude way of looking at this is that in many experiments, the more data you have, the better the chance that your finding is correct.[14] The presence of others allows for memories to be shared out and thus more accurately recalled. This is then further enhanced by culture. The norms we generate as a group of similar beings with shared values impinges on what our environment looks like. In turn, this shapes our internal dynamics by our vital co-existence with it. Like we see in Bartlett's story of the *War of the Ghosts*, our own idea on what a story should be like impinge on our ability to recall it correctly. We can be utterly convinced of our remembrance of the story, but be completely wrong because we are only able to reach a dynamic states that fit within the norms culture provides.

Added to all this is written language and symbols. This is a store in the more strong sense, as we can record things in ways which are both true and accurate. Our interpretation is still shaped by our own internal dynamics, but the information is hard-coded out there in the world to be referred to as and when we need it. Writing provides permanent fixed attractor states for our memory system to latch onto[15]. Here we come full circle. It is this capacity to hard-code information that has led us for many years to theorise cognition as symbolic, as storing and representing the world accurately and truthfully.

III.V Summarising EnM

We are now in a position to summarise. To address this in a systematic manner, we shall break this up into parts commonly associated with memory research: Encoding, Consolidation, Retrieval and Forgetting.

---

[14] Here too is a prime example of where we assume that correct = true. Of course, probability values say that it is likely that given the data there is a difference or similarity, but makes no claims on truth. Just that the data is quantitatively distinct.

[15] John Sutton has spent a long time investigating this idea under the banner of 'transactive memory'. Our books, social groups and cultures all help to distribute our cognitive processes outside of the skull (i.e. an extended mind). Sutton avoids using enactive, embodied or other paradigms largely to better allow for empirical research to be easily incorporated into the theory (Michaelian & Sutton, 2013). I agree with his position, but feel we should not shy away from bringing the benefits of **AE** along too.

Encoding Our system requires at least some amount of information to be passed into it that applies to the current experience. What we encode is the pattern of activity by maintaining elements of it as traces which dispose the system to return to this state when close to it in future. This is through creating attractor regions in phase space, which can influence the future states of the organism. How this is implemented is debatable, but is likely within the place cell/**VCO** populations of the hippocampus and **EC**.

Consolidation Structural and biochemical changes at the level of synapse are necessary if these traces are to be maintained at least in the short-term. These changes are non-specific to a memory however, and are utilised to retain the dispositional properties of the state as a memory trace. These properties are not confined to any one cell, but across a network of linked cell populations

Retrieval Subsequently, retrieval is the reactivation of the network which produces the same state, or something close to it. The extent to which this reactivation of the network can be said to match to the previous one is only possible by an outside observer. Furthermore, this return to a previous state will be subject to change precisely because the environmental interactions are unlikely to be the same. When victims of crime are asked to return to the scene, when we go back into a room to 'jog' our memory of something, we do so because it increases the chance of returning to the previous state accurately. We already understand the need during recall to be within a similar environment even without realising consciously.

Our memories are also active in their running. It is very difficult to recall aspects of an event immediately. Often, the scene must be re-presented in order (even if not it real-time) to be able to grasp the relevant information desired. When asked in court, eyewitnesses are often asked to 'talk through the events', as a way of ensuring information is returned that is less likely to be forgotten or lost as a result of cherry picking of times and sub-events within the memory. We move back into rooms in order to 'jog' our memory. When asked to recite a song I know well, I could do it reasonably well. Ask me to do it backwards, and I'd likely find it impossible. The information we enact when we recall is almost always directional, and requires the pattern of activity to run the same as it did initially.

Semanticisation removes some of the spatial specificity, doing so through some mechanism of attuning the state to predominantly internal ones. It is not a complete solution, but one which has enabled our memories to become portable even without the extra support of language. Stripped of most of their spatio-temporal bearings, these memories are enactive only in the sense that they arise only in the interaction of brain and body.

But it is important to stress that at no point in the reactivation itself is there any notion of content. This is applied by deliberative systems in the prefrontal cortex. The return to a network state enables the recalculation of hypotheses, with the most likely one being what is recalled. Our memories are so malleable precisely because hypothesis testing is the main reason behind their existence. If our priors are altered, so too is the scenario chosen to be regarded as the experienced one. We as subjective agents impose the meaning and reliability conditions upon the memory, which ultimately lacks discrete content.

Forgetting Without the ability to forget, we would all be in a similar position to S. His ability to recall things without mistakes came at the cost of being unable to use any of this information effectively. Every experience, every small fact or detail brought forth a wealth of related information that it was quite impossible for him to follow the original thread of stories without intense concentration. Internally representing everything that ever occurred left little capacity for abstract thought, resulting in the inability to truly learn things. Changes to the world are stored as separate things, as opposed to seeing the patterns that might underpin them. If memory was to operate in this way all the time, it would cease to be an autopoietic system, as there is no reciprocation between it and the environment.

III.VI Fixing up AE

When working with an ongoing theory, there are bound to be places where disagreement seeps in. Despite how well **AE** can help explain our memory system, there are a number of problems that I will acknowledge here.

The first concerns the impact of society on our cognition. Examples in **AE** literature often make use of the single bacterium making sense of the sucrose gradients it finds itself in. Society, and the effects of social interaction, is left to examples pertaining to human cognition specifically. This is a mistake. Bacteria depend on social interaction to survive, above and beyond mere proximity (Blanchard & Lu, 2015). If we are to still claim deep continuity between mind and life, then we need to be aware of this facet of life. Thompson (2011) has commented upon this before, noting it to be something to keep in mind when discussing **AE**. For now, further work is necessary to fully understand how to incorporate this into the framework.

The second worry alludes to the adopting of Jonasian anthropocentrism within **AE.** The privileged position of humans in **AE** texts precludes understanding cognition outside of

humans because it leads to "an anthropomorphic conception of other living organisms (De Jesus, 2016, p.134). This is indeed a major issue if we are to be speaking about memory from the ground up. Taking an approach which focusses on the human phenomenology will find itself at odds with the empirical work it wishes to utilise to make its case. This is tied into my worry earlier on Varela's conception of neurophenomenology- right now there is little evidence of a way of conceptualising phenomenology outwith humans. De Jesus makes the point that this issue is both epistemological and ontological, with **AE** "[running] the risk of simply projecting the analogues of human experiences down the phylogenetic scale" (De Jesus, 2016, p.134).

To fix the issue, De Jesus suggests replacing Jonasian phenomenology with that of Jakob von Uexküll. In short, he proposes a non-mechanistic, biosemiotic approach that would allow **AE** to avoid these problems of anthropocentrism and social undervaluing. Like **AE**, Uexküll's idea of 'Umwelt' emphasises coupling between environment and organism in a way which is *meaningful*:

> The concept of an Umwelt highlights the world of the organism 'from the inside'; not only do organisms actively contribute to the construction of their own worlds, but also that these 'worlds' are in fact infused with unique meaning, signification and value for the organism. Living organisms do not encounter neutral objects when they interact with them only meaningful ones. The Umwelt thus emphasises that the organism has a unique and meaningful point of view on its world whereby things matter. (De Jesus, 2015, p.136)

The main difference is that Umwelt emphasises that any study needs to begin with the organism and how the environment is important to it, not through its lived experience. De Jesus' own Biosemiotic Enactivism (**BE**) takes Umwelt and combines it with biosemiotic ideals to create his own explanatory framework. Whilst seemingly addressing some of the issues above, semiotics is a field just as confused with how to explain representational elements as we find ourselves. Time will tell if adopting semiotics, Umwelt or both are better for understanding memory from a non-human perspective. For now, I direct the reader elsewhere for further information (De Jesus, 2016; Rodríguez, 2016)

III.VII Alternative Interpretations

I warned earlier about the theoretical assumptions of empirical research, and how they potentially cause a sticking point in interpreting the data in a way outwith of how it was conceived. Even if we take the position that such worries are not as important in light of the amount of evidence which converges towards the **EnM** explanation, it is best to address this directly by providing alternative accounts. Here I intend to give a space for rebuttal of **EnM**,

and see if it can be fitted within the established frameworks without losing any explanatory power.

III.VII.i Cognitivist Accounts of EnM

Place cells and their siblings are quite readily amenable to cognitivism. Each cell type stands in for one or more features of the external environment the organism finds itself in. Our ability to navigate lies in the reading of the activity of these cells and combing them to provide a fully internal map. This map is not veridical, but representative of key elements of the environment and the animal's current location within it.

This map is updated through incoming sensory input, which is processed within the hippocampus and integrated into the navigational strategy if deemed advantageous to how to locate current goals. The elements of this map deemed important long-term are then transferred to neocortical storage during rest and sleep. This is done via **SWR** events generated in replay and preplay situations, and constitutes a storing of previous experience. When recalled, this information is reactivated, which is not to say it is remembered veridically, but that the representation is largely recalled accurately. Any errors are failures in encoding, transfer, or retrieval mechanisms, as a result of decay in synaptic strengths or signalling pathways.

This accounts for our ability to manipulate memories using optogenetic techniques (e.g. Redondo *et al.,* 2014). Each memory we reactivate is from a specific space/time point, and generates the appropriate response from the animal. These memories are totally independent of the external environment the animal is *currently* in, and may even be independent of the animal. The information, as it appears to be, could be transferred to another animal into the same type of cell network. In this account, as all we are dealing with is pure information, this memory transfer should result in the animal responding and experiencing something it has never previously been exposed to.

The cognitivist may also appeal to the other aspect of memory we have ignored. Procedural memory is the sum total of our habits, patterns of behaviour and stores of fine-grained actions such as writing or riding a bike. If we take this side as being one of patterns and habits, then we can still suggest that the declarative system possesses memories as symbols. The place and grid cells do more than covary with the environment, they are causally related and encode representations of that space as an internally held map. The directedness and interaction with the environment we place central to **EnM** is just the procedural memory system at work. Our hippocampus processes incoming stimuli and generates potential actions which are then

carried out by the motor system. These are then assessed by updating sensory input and comparing it to long-term semantic memories.

Petri and Mishkin (1994) have attempted to argue something similar. They argue that a dual systems model allows for behaviourist (habits, noncognitive and procedural) and cognitivist (memory, cognitive and declarative) to work in harmony. The habit system learns the probability of a stimulus generating a response. This gives the animal an inaccessible-to-consciousness disposition to act in a certain way for certain sensory inputs. The memory system stores neural representations of information which are later associated with others to create banks of knowledge. These two systems are to be anatomically distinct, and to effectively operate independent to one another.

If such a system is correct, then it should be possible to implant memories from one animal to another. As long as the capacity to move is the same, the memory element should be nothing more than structural changes to certain cellular networks such that the firing patterns match that in the original animal. The technology currently exists for such an experiment (see above), and there are attempts right now to do so (Redondo, personal communication). A distinct declarative system of this type would hard to explain in any way other than contentful symbolic states being extant in the neural system. Engrams and memory traces would be exactly as they are commonly thought to be. This is the realms of science fiction, and is unlikely to be the case. But I would not like to completely rule it out, especially given evidence of invariant cell networks for specific phenomena.

III.VII.ii Connectionist/Dynamical Accounts of EnM

Connectionist **EnM** adopts a closer position to vanilla **EnM** by retaining the dynamical approach. The major difference is the role of representations as explanatory devices in the connectionist account. This poses a direct challenge to **EnM**. After all, we have stated that all of this work to some extent automatically presumes them to be representational devices. A connectionist approach can build on this work and provide a modern account of memory which is not committed to any radical positions on organising systems and rejecting internal information processing that does not rely on environmental factors once inside the safety of the brain.

A dynamical account of **EnM** can go even further. There is no need for explicit representations, as in **EnM**, and a greater focus on the looping between organism and environment. In fact, it is only differentiated by the lack of autopoietic theory within it. So what is lost when we strip autopoiesis from **EnM**? As we mentioned earlier when defining **AE**,

the key loss is the interactional asymmetry. A dynamical account of memory gives no preference to the organism. Instead a memory could be generated and driven by the environment without any input from the agent aside from the interactions themselves. To claim that we may acquire memories in this way from our environment loses the fact that memory, unlike perception or action, is a higher order process than these. There is to an extent a necessary level of asymmetry needed for the system to function correctly, as otherwise the memories generated would benefit the organism only by chance. We do not remember everything because we have directedness toward the features of the environment which are relevant.

A potential refutation to this lies in the definition for cognition in enactivism itself, Remember that there we described it as being about the viability of the action, not the optimality. This works well to go against the 'genocentric' attitudes by Dennett (2007) and others. Yet here we may be asked the question as whether this claim also traps us when we say the lack of interactional symmetry is bad for a **DST** memory. Viable actions will include those with no benefit to the organism, and therefore we are mistaken in finding a difference between **EnM** and **DST** memory accounts. This only works if we ignore that the organism co-creates the environment it is in- viable actions are constrained by those which fit within the bounds of this environment, and ones which do not threaten the autonomy of the system.

Autopoietic theories outwith enactivism may also say the same thing, that to privilege the organism in the interaction is to make too stronger claims on the importance on one side of the relationship. Here the argument for preferring autopoiesis over **AE** reaches back to the idea of teleology and mechanistic tendencies. Autopoiesis can make a good claim that by adopting wholly mechanistic concepts it fits extremely well with the empirical work we outline- indeed this is something **BE** tries to show (**Section III.VI**). I have sided with Varela and Thompson because I feel that teleological accounts are necessary when speaking of cognition, and will leave it to others to provide a 'pure' autopoietic account of **EnM**.

III.VII.iii A View From Neuroscience

Neuroscientific enquiry is often removed from the theoretical notions we have discussed at length here. Given the wealth of potential descriptions and explanations at play in spatial memory however there has been a shift in recent years to attempt to couch some of the empirical work into a cognitive framework.

Gyrogy Buzsáki has worked for a number of years on memory, specifically its consolidation and internal organisation. In recent papers, he outlines a neuroscience-based model which

aims to capture cognition as arising from action in a similar way to **EnM**. Interestingly, Buzsáki (2013) compares allocentric navigation to **SM**, and egocentric movement to **EM**. Semantic knowledge is relatively separate from where and when it was acquired, much like allocentric navigation. Similarly, **EM** requires context and subjectiveness to plan for future events, much as path integration is used to shortcut and quickly make sense of the environment around an organism.

> To support memory effectively, a neural system evolved for navigation must meet two more requirements. It must have the capacity to store large quantities of seemingly unrelated, or orthogonal, representations, and it must be able to self-generate temporally evolving cell assembly sequences. (Buzsáki & Moser, 2013)

These requirements have been addressed in our explanations elsewhere. Schemas are a way in which we can retain a bank of information in an associative network, and oscillatory activity generates sequences in a way which also illustrates the system's autopoietic organisation.

His model focusses upon the theta rhythmic activity observed in the cells of the **EC** and hippocampus. One caveat with this hypothesis is that in primates and bats theta is only intermittently seen, or not at all. Thus in higher mammals there cannot be the same kind of primacy attached to theta-modulated phenomena. This does not rule out the idea is sound, just that our capacity to learn has evolved such that it is much better at operating without the need to be driven by movement *per se*. Theta-driven cell populations are mixed with those that do not, and yet all cell-type populations we have described show more increased firing and stability when theta rhythm is present.

Buzsáki (2013) hypothesises that this may be because primates now drive these cells primarily by visual means (i.e. saccades), such that they can be maintained even during immobility. If so, we can point to Noe's notion of sensorimotor affordances as a way in which this may fit within enactivist thinking. Our memories in the short term appear to rely on the environment to be maintained- we generate and solidify patterns of activity because we know where to look to acquire the right information out there in the world. If our **CM** was internally rich in content from the start, then we should find ourselves with little need for cells driven by visual cuing or otherwise. The initial set of experiences would be sufficient to maintain the pattern long enough to enter long-term memory.

Of course such a model is not without some commitments we have attempted to show are not required. Most notably is the requirement of representations and relatively rich models. For example, Buzsáki and Moser propose that the expanded brains of mammals over insects is due

to the need to store more representations of things, such that categorisation naturally emerges as a result:

> The growth of networks that enable the storage of millions of situations in the mammalian brain and the evaluation of the relationships among them may also form the basis for representing and categorizing explicit knowledge. The same mechanisms that define unique positions and their relationships in a map can be used to define or symbolize events, objects and living things. (Buzsáki & Moser, 2013, p.132)

Whilst a fully enactive account of **SM** has lain outside the scope of this paper, it is clear that the enactivising it is not impossible. The authors note that our best models of semantic knowledge and learning utilise topological spaces with similarity defined in a multi-dimensional vector space. This is identical to the current **CM** models seen earlier, and given the anatomical similarity present between the two systems, it seems highly likely the two are operating on the same neural mechanism.

The representative commitments made in these papers is of neuroreps, not mindreps. As such, their existence does underlie the data, but not in a way which denies us the ability to use it for our purposes. A lack of any strongly held theoretical constraints is showing itself in how data is reported on place cells and memory as our understanding of the mechanisms at play increases. In a later paper by Buzsáki we have a model which speaks of action and re-evaluating cognition to be more to do with anticipation and prediction:

> Regarding the self-organized brain activity as the fundamental brain operation and viewing the brain's responses to external perturbations as secondary actions offers a new paradigm, in which "*meaning*" is acquired by matching preexisting neuronal patterns to action-perception (Buzsáki et al., 2014. p.48)

The authors echo the potential that mental time travel may occur in a real-time way, and thus be akin to, if not essentially *be*, a form of navigatory activity. Work such as this begins to alter the methodology of neuroscience such that enactivist and embodied positions may be better understood without prior theoretical constraints. As a theoretical framework, embodied neuroscience seems likely, with enactive neuroscience not that far behind.

IV. Rounding Up

IV.I Memory Revolution?

In reality, this work does nothing more than bring to light a position which has existed for some time. Couched in the terminology of our cultural definition of memory it makes little sense, and indeed runs counterintuitively to our thinking. But in the experiments carried out,

in the theories generated, the notion of memory as some representative machine has become less and less visible.

Instead we are faced with the facts that we have never had any really strong grasp on what memory *is*. Definitions and explanations have revolved around paradigms generated in other disciplines, with memory taking on whatever guise is most fashionable. Under this however is a system which has always been vaguely understood as there, but not seen as fundamental as it now appears. Our revolutionary idea has been there all along in plain sight.

As the science and philosophy of memory has advanced, so too has the ability to avoid assuming the primary dominating feature to be storage of actual events in a rich contentful manner. Declarative memories are plastic, diffuse entities within cell populations which convey the general idea of an event by resurrecting attractor ruins from dispositional properties within the system. Understanding this works better if we are not committed to theoretical constraints around representation. Non-representation, whilst still being theoretical, is less rigidly fixed in its positions to allow for better understanding of what is going on in cognition.

IV.II Representational without being Representational

This account has not attempted to make grand statements on the representational debate. Instead, we have committed to a neuroscientific representational position because these are implicit in any data we have used. Yet in accepting these we do not necessarily accept the baggage that comes with representations within philosophy of mind. What we are committing to is that any description or initial explanation of complex mental systems requires representations in order to be coherently understood on the page. Eschewing them altogether runs the risk of clouding the importance of any new insights, as it focusses more on the representations than the argument.

Instead, I have tried to show how the representational 'job' that place cells, memory traces etc. are thought to do is nothing more than point towards the desired goal state. They are indicators for the system at different levels (both global and local) so as to allow for the arrival at system states which are relevant for the current interaction with the environment. They are representative only at the level at which we observe them to be when looking into the system and discretising its processes. Alone, the information supposedly carried is no more representative than noise unless you know where to look. And the only way we know where to look is because of how the system is at a given time.

Enactivism is still very much a view which is often amenable to radical departures away from traditional representational stances. But it is not one which has to renounce them entirely in order to exist. This 'soft' enactivism still is radical in how it approaches mental states, it just does not find the need to couch the difference within the sphere of representational attitudes. A minimal form of representation exists as it were in the idea of dispositional properties being held, as well as within the **SM** debate. I feel I have shown that this representation is quite different from that commonly attributed to this field, but I stop short at saying this way of conceptualising is the *only* correct one.

IV.III Enabling Enactive Neuroscience

Our findings are quite consistent with a growing trend in memory science towards looking outside the confines of the skull to understand what's going on inside. Yet this way of thinking cannot succeed if it retains even slightly any cognitivist notions. Instead, our descriptions and explanations must retain a sense of the individual organism, and avoid the temptation to make claims of universal phenomena too hastily.

I want to propose a truly enactive neuroscience will provide answers to questions long felt to be unanswerable or intractable. We feel the need to stay rigidly to science as objectivity even when talking of the very things that enable us to know what that means. We as individuals learn about the world in different ways and enact different environments. To ignore the individual in all this removes more than just uncertainty from subjective positions. It is to leave out that cognition is about existing and surviving within the world for the organism. The mechanisms of cognition may be shared, but these mechanisms are concerned with the organism as individual. Take it out of this context and it has no purpose, and no way of expressing this cognition.

When we reconstruct data from a brain, ascribing it as a representation and internalised content, we only can do so because we observe it with a system geared towards using this data. The data itself only indicates what is happening when placed within a greater context; reconstructing the movements of a rat requires knowledge of the environment the rat was in and someone to separate out the signal into cells. Yet as we have seen, this data can be manipulated such that it does not represent the world objectively, but as the organism thinks it to be. Which for the purposes of acting in the world, is enough.

Enabling this position will first require the baby step into embodied cognition. There the notions of representation and content are less strongly challenged, whilst still allowing for

neuroreps to shift further in meaning from mindreps. There is already evidence this is occurring. One very recent example has been the immune system and gut flora. In these cases each has been shown to play roles in social behaviour and individual personality traits which far surpasses some indirect signalling between systems and/or organisms. Instead, there is a pronounced and dynamic coupling within which interactions are critical for the major cognitive agent to function as 'normal' (Gacias *et al.*, 2016). There has also been an appeal to treat emotions in a manner similar to how we have treated memory here (Kiverstein & Miller, 2015). The respect that these ideas now hold showcase a change in how we look at cognition in philosophy, informatics and neuroscience is beginning to take root.

V. Conclusions

Over the course of this paper I have attempted to sketch out a core notion of declarative memory rooted in enactivism. Our natural memory system is emergent from the complexities of our navigational capacity, and supplies us with the ability to plan ahead based upon previous experiences. This planning however does not necessitate the need for any veridical store of prior experience, just that we retain the gist of what happened (or could have happened). Natural memory acts as a mechanism by which we mentally time travel, testing hypotheses and supplying us with the necessary confidence to act in the world so as to enact these tested hypotheses in the environment proper.

This notion of enacting by weighting the potential of future actions ties in neatly with the autopoietic ideals of Varela, Thompson and others. Our organisation as organisms self-generates these systems of directedness into the world by the interactional asymmetry between us and the environment we co-create. Memory is not an objective take on the world, nor merely a subjective store of experience within an environment. It is a dynamic, directed hypothesis testing system that emerges from our need to navigate and make sense of the environment we help create. This places it not just within our own environmental sphere, but of those we interact with and our cultural history. Taking the 'middle way' improves our understanding of what memory is without recourse to contradictory statements about how best to categorise it. Because of this (amongst other things) **AE** is best poised to take our insights into the empirical data gathered thus far much further than has been assumed.

Of course, this preliminary work cannot quite rid the spectre of representation from enactive memory- the idea is too entrenched, too deeply assumed within the work used here. What I have shown is that this is relatively unimportant for our purpose, and can be left to others to decide if this poses any future problems. For now, I have suggested a minimally representative

core memory system is reinforced by the presence of artificial memory, which is enabled by access to language and the co-opting of perception into internal visualisations. These provide a representational flavour to memory, perhaps supplying these enactions with content. **SM**, whilst trickier to explain, fits within this framework also, but is beyond my reach here.

Alternative explanations of the data are feasible, and not completely refutable in the sketch laid out. Dynamical memory rooted in connectionist and **DST** theory provides an almost identical initial explanation of the work as **AE** does. What differs is in the commitment to what neuroreps are and their importance in shaping research and understanding. Dynamic memory will still improve our understanding of cognition and remove some of the misconceptions we have of its functions. But it misses the chance to revolutionise how we go about carrying out research on what is more than an objectively measurable system.

Our ability to remember is critical to us being able to refer to ourselves as a continuous entity. Yet this function is one of serendipity- it arose not with this purpose in mind, but from the increasing complexity of autopoietic systems that draw ever more fine grained directedness towards the complex environments they exist within. An organism more aware of itself in relation to its environment has resulted in an explosion in the capacity to retain knowledge in external stores, as well as to do so within the confines of the organism. But, as we have shown this is all secondary to why the system exists. To understand natural memory, working from the ground up is I believe the best way of tackling it, with **AE** providing the best way of doing so.

Bibliography

Allen, T. A., & Fortin, N. J. (2013). The evolution of episodic memory. *Proceedings of the National Academy of Sciences*, *110*(Supplement 2), 10379-10386.

Anderson, M. I., & Jeffery, K. J. (2003). Heterogeneous modulation of place cell firing by changes in context. *The Journal of Neuroscience*, *23*(26), 8827-8835.

Babichev, A., Cheng, S., & Dabaghian, Y. A. (2016). Topological schemas of cognitive maps and spatial learning. *Frontiers in computational neuroscience*, *10*.

Barandiaran, X. E. (2016). Autonomy and Enactivism: Towards a Theory of Sensorimotor Autonomous Agency. *Topoi*, 1-22.

Bartlett, F. C. (1932). *Remembering: An experimental and social study. Cambridge: Cambridge University*.

Bechtel, W. (2016). Investigating neural representations: the tale of place cells. *Synthese*, *193*(5), 1287-1321.

Bennett, A. T. (1996). Do animals have cognitive maps? *Journal of Experimental Biology*, *199*(1), 219-224.

Blanchard, A. E., & Lu, T. (2015). Bacterial social interactions drive the emergence of differential spatial colony structures. *BMC systems biology*, *9*(1), 59.

Brandon, M. P., Koenig, J., Leutgeb, J. K., & Leutgeb, S. (2014). New and distinct hippocampal place codes are generated in a new environment during septal inactivation. *Neuron*, *82*(4), 789-796.

Brito, C. F., & Marques, V. X. (2016) Is there a role for computation in the enactive paradigm? In *Fundamental Issues of Artificial Intelligence.* Müller, V. C. (Ed.).

Brown, R., & Kulik, J. (1977). Flashbulb memories. *Cognition*, *5*(1), 73-99.

Brown, G. D., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological review*, *114*(3), 539.

Burton, A. M., Bruce, V., & Hancock, P. J. (1999). From pixels to people: A model of familiar face recognition. *Cognitive Science*, *23*(1), 1-31.

Budson, A. E., Sullivan, A. L., Daffner, K. R., & Schacter, D. L. (2003). Semantic versus phonological false recognition in aging and Alzheimer's disease. *Brain and cognition, 51*(3), 251-261.

Burgess, N., & O'Keefe, J. (1996). Neuronal computations underlying the firing of place cells and their role in navigation. *Hippocampus, 6*(6), 749-762.

Buzsáki, G. (2015). Hippocampal sharp wave-ripple: A cognitive biomarker for episodic memory and planning. *Hippocampus*, *25*(10), 1073-1188.

Buzsáki, G., & Moser, E. I. (2013). Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nature neuroscience*, *16*(2), 130-138.

Buzsáki, G., Peyrache, A., & Kubie, J. (2014). Emergence of cognition from action. In *Cold Spring Harbor symposia on quantitative biology* (Vol. 79, pp. 41-50). Cold Spring Harbor Laboratory Press.

Carruthers, P. (2002). The cognitive functions of language. *Behavioral and Brain Sciences*, *25*(06), 657-674.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(03), 181-204.

Clark, A. (2015). Radical predictive processing. *The Southern Journal of Philosophy*, *53*(S1), 3-27.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*(1), 7-19.

Cheng, S., Werning, M., & Suddendorf, T. (2016). Dissociating memory traces and scenario construction in mental time travel. *Neuroscience & Biobehavioral Reviews*, *60*, 82-89.

Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, *23*(2), 157-205.

Christoff, K., Keramatian, K., Gordon, A. M., Smith, R., & Mädler, B. (2009). Prefrontal organization of cognitive control according to levels of abstraction. *Brain research*, *1286*, 94-105.

Clapin, H., Staines, P., & Slezak, P. (Eds.). (2004). *Representation in mind: new approaches to mental representation* (Vol. 1). Elsevier.

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

De Brigard, F. (2014b). The nature of memory traces. *Philosophy Compass*, *9*(6), 402-414.

De Brigard, F., Addis, D. R., Ford, J. H., Schacter, D. L., & Giovanello, K. S. (2013). Remembering what could have happened: Neural correlates of episodic counterfactual thinking. *Neuropsychologia, 51*(12), 2401-2414.

De Jesus, P. (2016). From enactive phenomenology to biosemiotic enactivism. *Adaptive Behavior*, 1059712316636437.

Dennett, D. C. (2011). Shall we tango? No, but thanks for asking. *Journal of Consciousness Studies*, *18*(5-6), 23-34.

Di Paolo, E. A. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the cognitive sciences*, *4*(4), 429-452.

Dragoi, G., & Tonegawa, S. (2011). Preplay of future place cell sequences by hippocampal cellular assemblies. *Nature*, *469*(7330), 397-401.

Dudai, Y. (2012). The restless engram: consolidations never end. *Annual review of neuroscience*, *35*, 227-247.

Eichenbaum, H., Sauvage, M., Fortin, N., Komorowski, R., & Lipton, P. (2012). Towards a functional organization of episodic memory in the medial temporal lobe. Neuroscience & Biobehavioral Reviews, 36(7), 1597-1608.

Evans, T., Bicanski, A., Bush, D., & Burgess, N. (2016). How environment and self-motion combine in neural representations of space. *The Journal of physiology*.

Foglia, L., & Grush, R. (2011). The Limitations of a Purely Enactive (Non-Representational) Account of Imagery. *Journal of Consciousness Studies*, *18*(5-6), 35-43.

Fodor, J. A. (1975). *The language of thought*. Harvard University Press.

Foster, T., Castro, C. A., & McNaughton, B. L. (1989). Spatial selectivity of rat hippocampal neurons: dependence on preparedness for movement. *Cell, 6*, 4.

Friston, K. J., & Price, C. J. (2003). Degeneracy and redundancy in cognitive anatomy. *Trends in cognitive sciences, 7*(4), 151-152.

Gacias, M., Gaspari, S., Santos, P. M. G., Tamburini, S., Andrade, M., Zhang, F., ... & Zachariou, V. (2016). Microbiota-driven transcriptional changes in prefrontal cortex override genetic differences in social behavior. *eLife, 5*, e13442.

Hafting, T., Fyhn, M., Molden, S., Moser, M. B., & Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature, 436*(7052), 801-806.

Harter, D., & Kozma, R. (2005). Chaotic neurodynamics for autonomous agents. *IEEE Transactions on Neural Networks, 16*(3), 565-579.

Hutto, D. D. (2011). Philosophy of mind's new lease on life: Autopoietic enactivism meets teleosemiotics. *Journal of Consciousness Studies, 18*(5-6), 44-64.

Hutto, D. D. (2016). Remembering without Stored Contents: A Philosophical Reflection on Memory. In *Memory in the Twenty-First Century* (pp. 229-236). Palgrave Macmillan UK.

Hutto, D. D., & Myin, E. (2012). *Radicalizing enactivism: Basic minds without content*. MIT Press.

Infeld, E., & Galkowski, A. (1997). *Nonlinear Dynamics, Chaotic and Complex Systems: Proceedings of an International Conference Held in Zakopane, Poland, November 7-12 1995, Plenary Invited Lectures*. Cambridge University Press.

Ishikawa, M., Doya, K., Miyamoto, H., & Yamakawa, T. (Eds.). (2008). *Neural Information Processing: 14th International Conference, ICONIP 2007, Kitakyushu, Japan, November 13-16, 2007, Revised Selected Papers* (Vol. 4984). Springer.

Izhikevich, E. M. (2007). *Dynamical systems in neuroscience*. MIT press.

Jacobs, J. (2014). Hippocampal theta oscillations are slower in humans than in rodents: implications for models of spatial navigation and memory. *Phil. Trans. R. Soc. B, 369*(1635), 20130304.

Josselyn, S. A., Köhler, S., & Frankland, P. W. (2015). Finding the engram. *Nature Reviews Neuroscience, 16*(9), 521-534.

Kay, K., Sosa, M., Chung, J. E., Karlsson, M. P., Larkin, M. C., & Frank, L. M. (2016). A hippocampal network for spatial coding during immobility and sleep. *Nature, 531*(7593), 185-190.

Kitamura, T., Pignatelli, M., Suh, J., Kohara, K., Yoshiki, A., Abe, K., & Tonegawa, S. (2014). Island cells control temporal association memory. *Science, 343*(6173), 896-901.

Kiverstein, J., & Miller, M. (2015). The embodied brain: towards a radical embodied cognitive neuroscience. *Frontiers in human neuroscience, 9*, 237.

Knierim, J. J., Kudrimoti, H. S., & McNaughton, B. L. (1995). Place cells, head direction cells, and the learning of landmark stability. *The Journal of Neuroscience, 15*(3), 1648-1659.

Koutstaal, W., Schacter, D. L., Galluccio, L., & Stofer, K. A. (1999). Reducing gist-based false recognition in older adults: encoding and retrieval manipulations. *Psychology and aging, 14*(2), 220.

Kropff, E., Carmichael, J. E., Moser, M. B., & Moser, E. I. (2015). Speed cells in the medial entorhinal cortex. *Nature.*

Kurth-Nelson, Z., Economides, M., Dolan, R. J., & Dayan, P. (2016). Fast Sequences of Non-spatial State Representations in Humans. *Neuron, 91*(1), 194-204.

Kyselo, M. (2014). The body social: an enactive approach to the self. *Frontiers in Psychology, 5*, 986, 1-16.

Langston, R. F., Ainge, J. A., Couey, J. J., Canto, C. B., Bjerknes, T. L., Witter, M. P., ... & Moser, M. B. (2010). Development of the spatial representation system in the rat. *Science, 328*(5985), 1576-1580.

Lashley K. S. (1950). In search of the engram. *Society of Experimental Biology Symposium No. 4: Physiological mechanisms in animal behaviour*. Cambridge, England, Cambridge University Press. 454-482.

Lavenex, P., & Amaral, D. G. (2000). Hippocampal-neocortical interaction: a hierarchy of associativity. *Hippocampus, 10*(4), 420-430.

Lee, I., Yoganarasimha, D., Rao, G., & Knierim, J. J. (2004). Comparison of population coherence of place cells in hippocampal subfields CA1 and CA3. *Nature, 430*(6998), 456-459.

Leutgeb, S., Leutgeb, J. K., Barnes, C. A., Moser, E. I., McNaughton, B. L., & Moser, M. B. (2005). Independent codes for spatial and episodic memory in hippocampal neuronal ensembles. *Science, 309*(5734), 619-623.

Lisman, J., & Redish, A. D. (2009). Prediction, sequences and the hippocampus. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 364*(1521), 1193-1201.

Luria, A. R. (1968). *The mind of a mnemonist: A little book about a vast memory*. Harvard University Press.

Luria, A. R. (1976). *Cognitive development: Its cultural and social foundations*. Harvard University Press.

Manns, J. R., & Eichenbaum, H. (2006). Evolution of declarative memory. *Hippocampus*, *16*(9), 795-808.

Marian, V., & Neisser, U. (2000). Language-dependent recall of autobiographical memories. *Journal of Experimental Psychology: General*, *129*(3), 361.

Marozzi, E., & Jeffery, K. J. (2012). Place, space and memory cells. *Current Biology*, *22*(22), R939-R942.

Maturana, H. R., & Varela, F. G. (1980). *Autopoiesis and Cognition: the Realization of the Living.* Berlin: Springer

Michaelian, K. (2011). Is memory a natural kind? *Memory Studies*, *4*(2), 170-189.

Michaelian, K. (2015). Opening the doors of memory: is declarative memory a natural kind? *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(6), 475-482.

Michaelian, K., & Sutton, J. (2013). Distributed cognition and memory research: History and current directions. *Review of philosophy and psychology, 4*(1), 1-24.

Miłkowski, M. (2015). The hard problem of content: Solved (long ago). *Studies in Logic, Grammar and Rhetoric*, *41*(1), 73-88.

Moser, M. B., Rowland, D. C., & Moser, E. I. (2015). Place cells, grid cells, and memory. *Cold Spring Harbor perspectives in biology*, *7*(2), a021808.

Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, *19*(3), 113-126.

Newen, A. (2011). Merits and Limits of a Philosophy of Autopoiesis. *Journal of Consciousness Studies*, *18*(5-6), 65-82.

O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain research*, *34*(1), 171-175.

O'Keefe, J., & Recce, M. L. (1993). Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus*, *3*(3), 317-330.

Orchard, J. (2015). Oscillator-interference models of path integration do not require theta oscillations. *Neural computation*.

O'Reilly, R. C. (2010). The what and how of prefrontal cortical organization. *Trends in neurosciences*, *33*(8), 355-361.

O'Regan, J. K., & Noë, A. (2001a). A sensorimotor account of vision and visual consciousness. *Behavioral and brain sciences*, *24*(05), 939-973.

O'Regan, J. K., & Noë, A. (2001b). What It Is like to See: A Sensorimotor Theory of Perceptual Experience. *Synthese*, 79-103.

Osvath, M., & Martin-Ordas, G. (2014). The future of future-oriented cognition in non-humans: theory and the empirical case of the great apes. *Phil. Trans. R. Soc. B*, *369*(1655), 20130486.

Oyama. S. (2011). Life in Mind: Commentary on Evan Thompson's "Mind in Life". *Journal of Consciousness Studies*, *18*(5-6), 83-93.

Petri, H. L., & Mishkin, M. (1994). Behaviorism, cognitivism and the neuropsychology of memory. *American Scientist*, *82*(1), 30-37.

Pfeiffer, B. E., & Foster, D. J. (2013). Hippocampal place-cell sequences depict future paths to remembered goals. *Nature*, *497*(7447), 74-79.

Plunkett, K., & Marchman, V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition*, *48*(1), 21-69.

Radvansky, G. A., & Copeland, D. E. (2006). Walking through doorways causes forgetting: Situation models and experienced space. *Memory & cognition*, *34*(5), 1150-1156.

Radvansky, G. A., Krawietz, S. A., & Tamplin, A. K. (2011). Walking through doorways causes forgetting: Further explorations. *The Quarterly Journal of Experimental Psychology*, *64*(8), 1632-1645.

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

Ramsey, W. (2014). Must cognition be representational? *Synthese*, 1-18.

Ramsey, W. (2016). Untangling two questions about mental representation. *New Ideas in Psychology*, *40*, 3-12.

Redondo, R. L., Kim, J., Arons, A. L., Ramirez, S., Liu, X., & Tonegawa, S. (2014). Bidirectional switch of the valence associated with a hippocampal contextual memory engram. *Nature*, *513*(7518), 426-430.

Rodríguez, S. (2016). Recurrences and Human Agential Meaning Grounding: Laying a Path in Walking. *Biosemiotics*, 1-16.

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.

Rucinska, Z. A. (2015). Pretence: role of representations and intersubjectivity? PhD Thesis.

Rudrauf, D., Lutz, A., Cosmelli, D., Lachaux, J. P., & Le Van Quyen, M. (2003). From autopoiesis to neurophenomenology: Francisco Varela's exploration of the biophysics of being. *Biological research*, *36*(1), 27-65.

Rumelhart, D. E., McClelland, J. L., & the PDP research group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Volume I.* Cambridge, MA: MIT Press.

Sargolini, F., Fyhn, M., Hafting, T., McNaughton, B. L., Witter, M. P., Moser, M. B., & Moser, E. I. (2006). Conjunctive representation of position, direction, and velocity in entorhinal cortex. *Science*, *312*(5774), 758-762.

Schlesiger, M. I., Cannova, C. C., Boublil, B. L., Hales, J. B., Mankin, E. A., Brandon, M. P., ... & Leutgeb, S. (2015). The medial entorhinal cortex is necessary for temporal organization of hippocampal neuronal activity. *Nature neuroscience, 18*(8), 1123-1132.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and brain sciences, 11*(01), 1-23.

Solstad, T., Moser, E. I., & Einevoll, G. T. (2006). From grid cells to place cells: a mathematical model. *Hippocampus, 16*(12), 1026-1031.

Solstad, T., Boccara, C. N., Kropff, E., Moser, M. B., & Moser, E. I. (2008). Representation of geometric borders in the entorhinal cortex. *Science*, *322*(5909), 1865-1868.

Squire, L. R. (1992). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. *Journal of cognitive neuroscience, 4*(3), 232-243.

Squire, L. R. (2009). The legacy of patient HM for neuroscience. *Neuron, 61*(1), 6-9.

Stensola, H., Stensola, T., Solstad, T., Frøland, K., Moser, M. B., & Moser, E. I. (2012). The entorhinal grid map is discretized. *Nature*, *492*(7427), 72-78.

Stuart, G., Spruston, N., Sakmann, B., & Häusser, M. (1997). Action potential initiation and backpropagation in neurons of the mammalian CNS. *Trends in neurosciences, 20*(3), 125-131.

Suddendorf, T., & Busby, J. (2003). Mental time travel in animals? *Trends in cognitive sciences*, *7*(9), 391-396.

Sutton, J. (2015). Remembering as public practice: wittgenstein, memory, and distributed cognitive ecologies. In *Mind, Language, and Action: Proceedings of the 36th International Wittgenstein Symposium* (pp. 409-443).

Talarico, J. M., & Rubin, D. C. (2003). Confidence, not consistency, characterizes flashbulb memories. *Psychological Science*, *14*(5), 455-461.

Takahashi, S. (2015). Episodic-like memory trace in awake replay of hippocampal place cell activity sequences. *Elife*, *4*, e08105.

Tang, Q., Burgalossi, A., Ebbesen, C. L., Ray, S., Naumann, R., Schmidt, H., ... & Brecht, M. (2014). Pyramidal and stellate cell specificity of grid and border representations in layer 2 of medial entorhinal cortex. *Neuron*, *84*(6), 1191-1197.

Taube, J. S. (2007). The head direction signal: origins and sensory-motor integration. *Annu. Rev. Neurosci.*, *30*, 181-207.

Thompson, E. (2005). Sensorimotor subjectivity and the enactive approach to experience. *Phenomenology and the cognitive sciences, 4*(4), 407-427.

Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind*. Harvard University Press.

Thompson, E. (2011). Reply to commentaries. *Journal of Consciousness Studies*, *18*(5-6), 176-223.

Thompson, E., Lutz, A. and Cosmelli, D. (2005). Neurophenomenology: An introduction for neurophilosophers. In A. Brook and K. Akins (eds), *Cognition and the Brain: The Philosophy and Neuroscience Movement.* Cambridge: Cambridge University Press.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological review*, *55*(4), 189.

Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., ... & Morris, R. G. (2007). Schemas and memory consolidation. *Science*, *316*(5821), 76-82.

Tse, D., Takeuchi, T., Kakeyama, M., Kajii, Y., Okuno, H., Tohyama, C., ... & Morris, R. G. (2011). Schema-dependent gene activation and memory encoding in neocortex. *Science*, *333*(6044), 891-895.

Tsuda, I. (2015). Chaotic itinerancy and its roles in cognitive neurodynamics. *Current opinion in neurobiology*, *31*, 67-71.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual review of psychology*, *53*(1), 1-25.

Van Cauter, T., Poucet, B., & Save, E. (2008). Unstable CA1 place cell representation in rats with entorhinal cortex lesions. *European journal of Neuroscience*, *27*(8), 1933-1946.

Van Gelder, T. (1995). What might cognition be, if not computation?. *The Journal of Philosophy*, *92*(7), 345-381.

Van Kesteren, M. T., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in neurosciences*, 35(4), 211-219.

Varela, F. J. (1996). Neurophenomenology: A methodological remedy for the hard problem. *Journal of consciousness studies*, *3*(4), 330-349.

Varela, F. J. (1999). The specious present: A neurophenomenology of time consciousness. *Naturalizing phenomenology: Issues in contemporary phenomenology and cognitive science*, *64*, 266-329.

Varela, F., Thompson, E., & Rosch, E. (1991). *The embodied mind.*

Villalobos, M., & Ward, D. (2015). Living systems: Autonomy, autopoiesis and enaction. *Philosophy & Technology*, *28*(2), 225-239.

Wallace, D. F. (2006). Roger Federer as Religious Experience. *The New York Times*, *20*.

Wang, S. H., & Morris, R. G. (2010). Hippocampal-neocortical interactions in memory formation, consolidation, and reconsolidation. *Annual review of psychology, 61*, 49-79.

Wargo, E. (2010, Jan 16) The Art of Memory: Why It's Just About the Coolest Thing Ever, and Why You Should Learn It Today. [Web log post]. Retrieved from http://thenightshirt.com/?page_id=186

Waters, J., Schaefer, A., & Sakmann, B. (2005). Backpropagating action potentials in neurones: measurement, mechanisms and potential functions. *Progress in biophysics and molecular biology*, *87*(1), 145-170.

Weber, A., & Varela, F. J. (2002). Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the cognitive sciences, 1*(2), 97-125.

Welday, A. C., Shlifer, I. G., Bloom, M. L., Zhang, K., & Blair, H. T. (2011). Cosine directional tuning of theta cell burst frequencies: evidence for spatial coding by oscillatory interference. *The Journal of neuroscience*, *31*(45), 16157-16176.

Welton, D. (2011). Can a Top-Down Phenomenology of Intentional Consciousness be Integrated with a Bottom-Up Phenomenology of Biological Systems? *Philosophy Today*, *55*(Supplement), 102-113.

Werning, M., & Cheng, S. (2014, July). Is Episodic Memory a Natural Kind?-A Defense of the Sequence Analysis. In *36th Annual Meeting of the Cognitive Science Society, Quebec City, Canada* (pp. 23-26).

Yartsev, M. M., Witter, M. P., & Ulanovsky, N. (2011). Grid cells without theta oscillations in the entorhinal cortex of bats. *Nature*, *479*(7371), 103-107.

Yates, F. (1966). *The art of memory*. Chicago: University of Chicago Press.

Zednik, C. (2011). The nature of dynamical explanation. *Philosophy of Science, 78*(2), 238-263.

Zhang, S., Schönfeld, F., Wiskott, L., & Manahan-Vaughan, D. (2014). Spatial representations of place cells in darkness are supported by path integration and border information. *Frontiers in behavioral neuroscience*, 8.

Zhang, S. J., Ye, J., Miao, C., Tsao, A., Cerniauskas, I., Ledergerber, D., ... & Moser, E. I. (2013). Optogenetic dissection of entorhinal-hippocampal functional connectivity. *Science*, *340*(6128), 1232627.