OXFORD

Space, Time, and Memory

edited by

Lynn Nadel and Sara Aronowitz

Space, Time, and Memory

Space, Time, and Memory

Edited by

Lynn Nadel
and
Sara Aronowitz

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data

Data available

Library of Congress Control Number: 2025931058

ISBN 9780192882547

DOI: 10.1093/oso/9780192882547.001.0001

Printed and bound by
CPI Group (UK) Ltd, Croydon, CR0 4YY

Contents

PART III

List of Contributors

Sara Aronowitz

Alexandria Boyle

Felipe De Brigard

György Buzsáki

G. William Chapman

Quan Do

EJ Green

Michael E. Hasselmo

Patrick A. LaChance

Samantha Malmberg

Ida Momennejad

Lynn Nadel

Nora S. Newcombe

Kim V. Nguyen

John O'Keefe

Mahir Patel

Ian Phillips

Charan Ranganath

Sarah Robins

Jennifer C. Robinson

János Végh

L. Kelton Wilmerding

Jacob H. Wilmot

Introduction

This volume reflects the outcome of a workshop on 'Memory, Space, and Time' held at the University of Arizona (UA) in November of 2022. The idea for the workshop came out of discussions between the two editors, both at the UA at that time. Both of us have worked on issues relating to space, time, and memory, though from rather different perspectives. This complementarity suggested an approach that we hoped would be fruitful, and we believe this volume confirms our initial enthusiasm.

Our plan was to invite speakers from three different disciplines—philosophy, psychology, and neuroscience. Maintaining these distinctions is of course challenging, given the interdisciplinary nature of science these days, but we hope the chapters in this volume convey both the breadth of thinking in the field and the ways in which combining these different perspectives can lead to significant advances in understanding how space, time, and memory interact in our minds/brains.

This volume is organized with such interdisciplinarity in mind, comprising three sections: the first, with chapters by Phillips, Nguyen & Newcombe, Buzsáki & Végh, O'Keefe, and Aronowitz & Nadel, focus on space and time as parts of memory. The second, with chapters by Robins, Ranganath, Hasselmo et al., and De Brigard, detail theories of memory in their own right, while the third, with chapters by Green, Boyle, and Momennejad, consider how learning, perception, and AI intersect with space, time, and memory.

We received support from a number of sources in mounting this workshop. NSF provided critical funding via a conference grant that enabled us to not only support our speakers but also to support the attendance of a diverse group of students (twelve in all) from both North America and Europe. We thank Betty Tuller at NSF for this support. We also received support from a number of sources at the University of Arizona, including the Cognitive Science Program, the Departments of Psychology and Philosophy, the Provost's Office and the VP for Research. Their support was crucial in making certain we could invite many local participants to join in both the workshop itself, and the social events associated with it.

The Oxford University Press agreed not only to publish this volume, but also to provide significant financial support for the workshop. We thank Martin Baum and his crew at OUP for their support. We are also indebted to April Chan and Aliya Dewey for help organizing the workshop, and Julia Minarik and Cameron Yetman for help putting together the volume.

Finally, we thank the participants for their contributions, both to an outstanding workshop and to this volume, which we hope readers will find stimulating.

Sara Aronowitz
Lynn Nadel

PART I

1

What Has Episodic Memory Got to Do with Space and Time?

Ian Phillips

Introduction

It is widely held that episodic memory is constitutively connected with space and time. In particular, many contend that episodic memory constitutively has spatial and/or temporal content: for instance, necessarily representing a spatial scene, or *when* a given event occurred, or at the very minimum that it occurred in the past.[1] Here, I critically assess such claims. I begin with some preparatory remarks on the nature of episodic memory. I then ask: How, if at all, is episodic memory constitutively spatial? And, how, if at all, is episodic memory constitutively temporal? In answer, I argue that episodic memory need not have any spatial content, nor (in any substantial sense) need it represent *when* its events occur, nor even that they occur in the past. Instead, only a relatively modest connection holds between episodic memory and time in virtue of the temporal structure of its objects. Finally, I critically assess whether considerations concerning the organization and encoding of episodic memory in creatures like us provide stronger reason to posit a constitutive link between our episodic memories and space or time.

What Is Episodic Memory?

Very broadly, we can distinguish two approaches to episodic memory. The first introduces episodic memory by way of a contrast. The second introduces episodic memory directly. Both approaches are found in Endel Tulving's classic paper, 'Episodic and semantic memory'. There, Tulving begins by noting Ross Quillian's (1966) introduction of the term 'semantic memory', before asking: 'What do we contrast with semantic memory?' Tulving then proposes that we 'refer to this other kind of memory, the one that semantic memory is not, as "episodic memory"' (1972:

[1] By a constitutive connection I mean a connection that cannot fail to hold (i.e., holds essentially) since it is part of what it is to be the kind in question (cf. Burge 2010: xiii–xv, Chs 1 and 3). Thus, to claim that episodic memory constitutively has spatial content is to claim that it is part of what it is for a state or episode to be (or count as) an episodic memory that it possesses spatial content. Such claims assume that the mind can be divided into kinds with distinct natures and that episodic memory is one such kind. I adopt this assumption in what follows.

Ian Phillips, *What Has Episodic Memory Got to Do with Space and Time?*. In: *Space, Time, and Memory*. Edited by: Lynn Nadel and Sara Aronowitz, Oxford University Press. DOI: 10.1093/oso/9780192882547.003.0001

384). Alan Baddeley similarly comments that early talk of episodic memory was typically contrastive: 'In its initial years … the term episodic memory was commonly used to refer to all [explicit] memory other than semantic or working memory' (2002: 5). Here, I raise a few questions about this *contrastivist* approach, on which the contrast with semantic memory takes primacy in determining the extension of the term 'episodic memory' (see Figure 1.1).

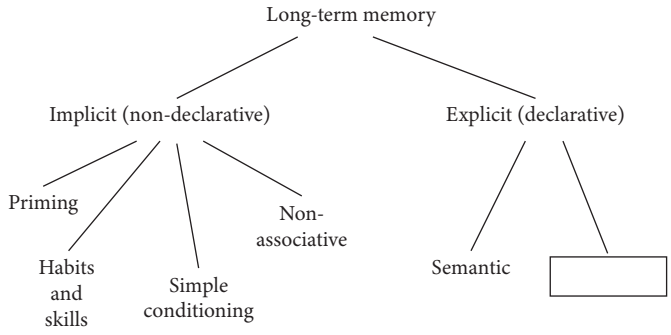


Figure 1.1 A standard taxonomy of the subdivisions of long-term memory based on Squire (1992), see likewise Hampton & Schwartz (2004). According to contrastivism, episodic memory is simply whatever memory belongs in the box on the right-hand side, i.e., explicit/declarative memory of a non-semantic kind.

A forceful recent recommendation of contrastivism can be found in Ali Boyle's paper, 'Remembering events and representing time'. There, Boyle suggests that the term 'episodic memory' was introduced to 'mark an exhaustive division in declarative memory' (2021: 2511). In line with this, she urges us to take the term 'to pick out the type of declarative memory that is not semantic' since 'marking this fundamental division in declarative memory is the central theoretical work the episodic/semantic distinction is supposed to do' (2513). As Boyle notes, David Rubin and Sharda Umanath adopt a similar stance towards what they call 'event memory' and which they take to be 'the fundamental natural kind that is an [sic] opposition to knowledge (i.e., semantic memory)' (2015: 2).

Several questions arise about this picture.

First, insofar as episodic memory is intended to 'mark an exhaustive division in declarative memory', we need to know what *declarative* or *explicit* memory is. This is no trivial question. Explicit or declarative memory might mean a form of memory which a person can *make explicit* or *declare* via a verbal statement. Such a characterization obviously cannot be accepted by anyone wishing to allow for the possibility of episodic memory in pre-linguistic infants and non-linguistic animals. Moreover, even theorists persuaded that episodic memory is unique to adult humans (e.g., Tulving and Markowitsch 1998) must justify according verbal reports so special a role—a role not standardly accorded to verbal reports in relation to other mental states, and one which would seem in tension with the fact that we often struggle to articulate our memories verbally. However, if we don't restrict 'making explicit' to

linguistic performances, more needs saying as to which performances are allowed to count—especially if the forms of memory depicted on the left of Figure 1.1 (e.g., habits and skills) are to be excluded since these evidently show up in behaviour.

Sometimes 'explicit' is understood in terms of consciousness. Few, however, would claim that we had a firm operational or theoretical grip on that notion (nor how exactly it applies to semantic memory). Another possibility is to understand explicit memory in terms of there being an internal representation of the memory's content in the system, as opposed either to the content being derivable from what is represented, or—more substantively—to its being reflected in rules wired into the system, and in that sense procedural. However, it is far from clear that this notion captures all the cases typically placed in the implicit (non-declarative) category in Figure 1.1 (e.g., semantic priming).

A more promising approach understands explicit memory in terms of representations which are available to a wide range of cognitive operations, such as belief fixation, planning, reasoning, acting, and reporting (cf. the concept of access consciousness articulated in Block 1995). Implicit memories would then be excluded insofar as the corresponding representations (if any) are only available to a narrow range of processes (e.g., perceptual or motor processes but not reasoning and reporting). Although promising, the proponent of this approach owes a more precise and principled account than the sketch just given. What constitutes a sufficiently wide range of processes? How do we handle the variability in such processes across species? And, what constitutes availability, exactly—actual access, potential access without further processing, or simply potential access even with further processing?

Let us bracket this first issue. A second, more immediate difficulty is whether we have sufficient grip on what *semantic* memory is, so as to define episodic memory in contrast to it, as *non-semantic* memory. One way to press this issue is to question whether semantic memory itself is a natural kind. Textbook examples of semantic memory involve factual knowledge such as the knowledge that Paris is the capital of France, or that Liz Truss was (briefly) UK Prime Minister (cf. Tulving 1972: 387). However, as Brown (ms) points out, plausibly many different kinds of memory have some claim to being semantic, either in being to some degree abstract, or involving a discursive representational format. Yet these various forms of memory lack any clear unity. In addition to standard textbook examples, consider: cognitive environmental maps, models of system dynamics in model-based reinforcement learning models, tacit grammatical knowledge, or the representations of what Susan Carey calls 'core cognition', and which she characterizes as a 'type of conceptual structure … that differs systematically from both sensory/perceptual representational systems and theoretical conceptual knowledge' in being richly conceptual yet created by innate perceptual analysers, domain specific, and iconic (2009: 10). These forms of memory do not obviously form a single natural kind or even genus. In turn, this casts doubt on the idea that explicit memory divides into *two* exhaustive kinds. Instead, it suggests a picture on which explicit memory comes in a variety of different forms, with various cross-cutting similarities and differences.

In addition to these memories, we should consider memories of objects, people, and places (e.g., my first bike, my grandmother, London in the 90s)—on which see Debus (2007) and Openshaw (2022). And, also, general (i.e., merged, summary, or prototype) event memories (e.g., visiting my grandparents as a child, reading stories to my daughter when she was little). Again, it is not obvious why we should feel forced to count these either as semantic or episodic, as opposed to distinct but related forms of explicit memory.[2] This final point connects to an issue highlighted in recent work by Andonovski (2020) and Aronowitz (forthcoming) namely that episodic memories are continually being transformed into *more* abstract, *more* 'semantic' memories via a suite of semanticization/schematization processes thought to correspond, at least roughly, to a transfer of encoding from hippocampal to cortical structures.

The key upshot is that we should be sceptical that episodic and semantic memory constitute two exhaustive natural kinds. On the face of it, 'explicit' memory comprises multiple distinct kinds of memory deriving from and/or contrasting with the prototypically episodic, with no obvious bipartite unity.

Where does this leave contrastivism? A helpful analogy here is with talk of the perception/cognition distinction. When people introduce such a distinction, presumably they do not intend an exhaustive division of the mind into two natural kinds, with the perceptual being understood as the non-cognitive, or vice versa. For one, someone interested in the perception/cognition divide might well happily recognize affective or motor systems which fall into neither category. For another, introducing such a distinction does not presuppose that it is a distinction between two natural kinds. Someone might, for instance, regard cognition as a ragbag of different kinds.[3] Instead, one plausible understanding is that they take perception to be an especially basic form of mind from which certain other forms derive (e.g., episodic memory, perceptual belief, etc.) and which can be fruitfully contrasted with other mental forms (e.g., abstract factual knowledge, propositional desire, intention, etc.).

Similarly, when we introduce episodic memory, we need not think of it as marking an exhaustive contrast with a singular contrasting kind, viz. 'semantic memory' but instead as highlighting a particularly basic form of memory from and with which certain other forms can be derived and contrasted. Such a proposal fits naturally with Brown's scepticism that semantic memory is a natural kind, and likewise with the idea that schematization processes produce 'a variety of intermediary forms of varying generality' (Andonovski 2020: 342) or 'range of contents differing in degrees of episodicity' (Aronowitz forthcoming).

[2] Compare Neisser (1981) on 'repisodic memory'. It is also not obvious how to place other forms of memory within the standard taxonomy, e.g., prospective memory, or short-term iconic, fragile visual and working memory. See further footnote 17 on observer memories.

[3] Compare Firestone and Scholl (2016: 1) who contrast perception with 'higher-level cognition … states such as beliefs, desires, emotions, motivations, intentions, and linguistic representations'.

This proposal of course behoves us to provide some *direct* characterization of episodic memory. What is this supposedly basic form of memory? Famously, Tulving offers a direct characterization of episodic memory, as memory of 'personally experienced unique episodes' (1972: 387) or 'memory for personal experiences and their temporal relations' (401–02)—a characterization he embellishes in subsequent work, invoking notions such as autonoetic consciousness and mental time travel (e.g., Tulving 2001; Wheeler, Stuss, and Tulving 1997). Tulving's original conception echoes Reid's pioneering discussion where he speaks of memory as involving the 'renewal of a former acquaintance with the thing remembered'—his example being his memory of the transit of Venus in 1769, which he says he 'must therefore have perceived … at the time it happened' (1785/2002: III.1, 253–55).[4]

Here, I propose that we adopt a kindred answer, on which episodic memory is memory *for personal experiences*. (For reasons to be discussed in §4, I omit Tulving's 'and their temporal relations' (1972: 402).) To unpack this idea, let us focus on Reid's sighting of the transit of Venus. Here, we have a perception of an event which takes place in 1769. Writing sixteen years later, Reid does not perceive the transit again. Instead, he has, in some sense, retained his past perception in memory. On the present proposal, this is understood in terms of his possession of a standing capacity to *represent* that particular perception together with its object (i.e., Venus' transit). The past perception and its object thus come to be before Reid's mind in virtue of a standing capacity which his 1769 perception has given rise to, and which can manifest in a specific episode of recall. Importantly, the transit comes to be before Reid's mind in 1785 in a quite different manner to the way it was before his mind in 1769. In 1769, it was the object of a present perception; in 1785 it is before his mind as the object of a represented past perception. Generalizing, episodic memory is the capacity to *representationally* return past experiential episodes to mind.[5]

On this view, episodic memory contrasts semantic memory with respect to its contents or objects. Certainly, we can have semantic memories of past events and experiences. But with semantic memory the relevant memory contents are *facts* or *propositions* about our pasts, with past occurrences at best figuring as constituents of these. The contents of episodic memory are not facts or propositions but *events* or *occurrences*.

Various concerns beset such a picture. Let me briefly mention three. The first is that the picture imposes sophisticated metacognitive requirements on episodic memory. However, although the present proposal claims that episodic memory involves the representation of one's prior perceptions, it does not claim that such

[4] For discussion of Reid's view, see Copenhaver 2006. On Copenhaver's interpretation, there are important differences between Reid's view and that here defended. First, whilst Reid takes the objects of memory to be events, he denies that these events are experiences, insisting that they are (at least in relevant cases) worldly not mental events. The view here claims that episodic memory has *both* perceptual experiences and their worldly objects as its objects: Reid remembers the transit of Venus by remembering perceiving it. Second, at least on Copenhaver's 'constitutive' interpretation, Reid takes a belief that the remembered event was a past occurrence to be an essential ingredient of memory. No such belief—nor even any tensed content—forms part of the present account.

[5] For a now classic elaboration of this picture, see Martin 2001. See also Martin 2015 and 2019.

representation is conceptual, requiring (e.g.) possession of the concept *perception*. Nor does it require any cognitive attitude, such as a belief that the recalled event occurred in one's past. Consequently, there is no obvious reason to think that the present picture is inconsistent with non-human animals who lack such concepts or attitudes having episodic memories. Compare how we might happily attribute perceptual object-representations to a non-human animal without requiring that they possess the concept *object*.

A second familiar concern is that the process of episodic recall involves a great deal of (re)construction and supplementation relative to impoverished traces. Many have suggested that this rules out thinking of episodic memory as genuinely renewing our former apprehension or acquaintance.[6] In reply, a perceptual analogy is helpful. In vision, too, there is notorious underdetermination of what we see by the proximal retinal image. In solving this underdetermination problem, a critical role is played by inferential processes exploiting background knowledge (e.g., Bayesian inference exploiting learned and/or innate perceptual priors).[7] This is certainly an impressive achievement. As Gregory famously puts it: 'We are given tiny distorted upside-down images in the eyes, and we see separate solid objects in surrounding space. From the patterns of stimulation on the retinas we perceive the world of objects … this is nothing short of a miracle' (1966: 7). Yet, however seemingly miraculous, few conclude that vision does not afford us a way of perceiving environmental particulars (though, see, e.g., Seth 2021). Instead, the relevant processes of supplementation and reconstruction based on stored information are taken to be part of what *allow* us to perceive our perceptual environments.[8] It is not obvious why a similar response cannot be given to the constructivist challenge in the case of episodic memory, namely that we should think of the processes of (re)construction and supplementation involved in retrieval as what allow us to retain and renew acquaintance on the basis of severely impoverished traces.[9]

A final concern is pressed by Aronowitz (forthcoming) who argues that retained acquaintance is not the *kind* of thing that can be semanticized. The precise issue here is delicate. Aronowitz presses it in light of two different (and neither uncontroversial) ways of conceptualizing semanticization, viz., Buzsáki and

[6] Schacter and Addis review relevant empirical work, arguing that memory errors 'provide critical evidence for the fundamental idea that memory is not a literal reproduction of the past, but rather is a constructive process in which bits and pieces of information from various sources are pulled together' (2007: 773). Such ideas trace back to Bartlett 1932. For philosophical discussion, see Michaelian 2012, 2016 and De Brigard 2014.

[7] This at least is the classic view tracing back to von Helmholtz 1910/1962, and elaborated in (e.g.) Gregory 1966, Marr 1982, Rock 1983, and Knill and Richards 1996. For an alternative perspective, see Gibson 1979, reviewed in Rogers 2021.

[8] Some argue that whilst constructivist models of perception are compatible with the perception of environmental particulars, they are incompatible with relational or naïve realist accounts of perception. For a reply on behalf of the naïve realist, see French and Phillips 2023, also Campbell 2002 and 2011. A related argument might be made against the present view of episodic memory; I suggest that an analogous reply is available.

[9] Of course, the perceptual analogy only goes so far and much more needs saying to provide a full defense of the present picture of episodic memory. For instance, as Matthew Soteriou pointed out to me, a significant difference between perception and memory is that in perception the question of which particular is being perceived is partly a matter of which particular is presently triggering and sustaining one's perceptual state. In contrast, memory is a standing capacity which can be exercised in the absence of its object. As a result, one might argue that sensory memory cannot alone secure reference to particular past episodes without the help of cognition.

Moser's (2003) navigational theory on which semanticization reflects computations analogous to those involved in shifting from egocentric to allocentric encodings of our spatial environments; and McClelland et al.'s (1995) complementary learning systems approach, on which semanticization involves extracting information from richly detailed spatiotemporal representations through a process of repeated offline 'replay' to produce abstract, general representations. On neither theory is it entirely clear what constraint is imposed by recognizing that episodic contents must be amenable to semanticization. Nor crucially is it clear that we should accept the implicit assumption in play that the categories and (presumably, representational) kinds of such theories neatly map onto the categories and kinds of personal level psychology (cf. the discussion of perceptual kinds in Phillips 2018; French and Phillips 2023; and Campbell 2011).

Again, however, it is worth noting that an analogous 'interface problem' arises in the case of perception. Perception seemingly effortlessly revises, creates and updates our beliefs. This undoubtedly poses an explanatory challenge. But it is far from obvious that it provides decisive reason to think that perception and cognition must exploit a common format. To the contrary, many theorists insist that it is precisely differences in format which help characterize the distinction between perception and cognition. For instance, Burge (2022) holds that natural, and so human, perceptual representation is constitutively iconic and non-propositional (e.g., 2022: 331), and Block (2023) makes the stronger claim that perception quite generally is constitutively iconic and non-propositional (2023: Chs 4–5). And whilst it is true that some theorists do argue in favour of architectural views and against format-based accounts on the grounds that the best explanation of the ease with which perception updates belief is that it outputs representations in the same discursive representational format as thought (e.g., Mandelbaum 2018; Quilty-Dunn 2022: 809), these theorists face their own 'interface problem'. This is because such theorists recognize a plurality of representational formats *within* perception. They must then explain how fast inferential processing can occur *within* perception despite such differences of format, in turn raising the question of why any such explanation cannot explain transitions between perception and cognition.

These considerations do not directly answer Aronowitz's challenge. But to the extent that one is sanguine that the analogous challenge can be met in relation to perception, they suggest that we should not be over hasty in concluding that semanticization is inconsistent with the picture of episodic memory here espoused.

Relatedly, and in line with what Andonovski calls the 'transitional gradation challenge' (2020: §4.2), one might press that the gradual nature of the transitions between forms of memory threatens a conception of episodic memory as retained acquaintance. But the existence of a continuum of hybrid cases involving purely episodic as well as more general elements (or similarly memories which summarize two, three, four … episodic memories) does not call into question the existence nor explanatory priority of purely episodic memories. In the same way, theorists like Block and Burge can, and readily do, acknowledge that there exist a range of

non-perceptual states which are immediately derived from and exploit or incorporate iconic and perceptual elements (see, e.g., Block 2022: 215 and Burge 2023: 332 on perceptual beliefs). This is consistent with the existence of purely iconic states.

With these preliminary remarks behind us, I now turn to the core questions of the paper: To what extents, if any, is episodic memory spatial and temporal?

Is Episodic Memory Constitutively Spatial?

To address the central question of this section, I take as my stalking horse the view proposed by David Rubin and Sharda Umanath in an important theoretical paper from 2015.[10] There, Rubin and Umanath distinguish between *event memory* and *semantic memory*. On their account, the key feature distinguishing event memories from semantic memories is their spatial content. Specifically, they tell us that 'an event memory is a mentally constructed scene', where a scene is 'an organized spatial layout that locates the person remembering relative to the rest of the scene' (Rubin 2022: 467–68). For Rubin and Umanath, episodic memories are a special subclass of event memories marked out by further features in which the represented events are unique, about the self, relived, and voluntarily constructed. However, being a subclass of event memories, episodic memories must also in their view have a distinctive spatial content.

Rubin and Umanath's view has been influential in philosophy. In particular, Boyle (2021) suggests that we endorse their view as a view of episodic memory in general, in other words, that we identify Rubin and Umanath's event memories with episodic memory—event memories being (as discussed above) the fundamental natural kind which contrasts semantic memory. Rubin and Umanath's view also has various Kantian precedents. Consider, for instance, Russell and Hanna's *Kantian Minimalism* according to which 'just as experience must involve an experiencer's spatial perspective on objects and events, so too must re-experience' (2012: 33).[11]

What argument do Rubin and Umanath give in favour of their picture? Here is one key passage:

> An event memory must have spatial organization; without it, the memory … will be judged as knowledge. … One would judge one's recall to be a memory of an event only if it was experienced as the recall of a [spatial] scene. (2015: 4)

Why think that it is true that without a spatial scene we would judge an event memory instead to be knowledge, i.e. a semantic memory? A clue comes in a second passage, where Rubin and Umanath offer the following consideration.

[10] Rubin and Umanath are far from alone amongst scientists in insisting on a tight connection between space and episodic memory. For instance, Moser et al. aver that 'space is a central element of all episodic … memories' (2015: 3).

[11] Russell and Hanna make the same claim about time, taking experience to be necessarily spatiotemporal, and therefore also in their view episodic memory conceived as 're-experience'.

> Because a scene cannot be imagined … without an assumed viewpoint, … the self enters as a locus in space and time … from which the scene is remembered. … This egocentric perspective from a specific spatial location is what distinguishes event memory from knowledge in phenomenological terms. (2015: 1)

How should we reconstruct this argument? The basic idea appears to be that there is a structural, phenomenological difference between event (or episodic) memory and knowledge (or semantic memory). This structural difference is held essentially to involve the presence of an egocentric perspective within the content of episodic memory. This perspective is then claimed to be a spatial viewpoint on the recalled scene. In turn, this warrants the conclusion that, to be distinguished from semantic memory, episodic memory must involve a spatial viewpoint on a spatial scene.

The most obvious concern with this argument is that, whilst it seems right to hold that a *spatial scene* cannot be episodically recalled except from a spatial viewpoint (and thus that episodic memories of *spatial scenes* must involve a spatial viewpoint), no reason is given for thinking that episodic memories *in general* must involve a spatial viewpoint. To assume that all episodic memories are of spatial scenes would simply beg the question.[12]

Let us then grant that episodic memory critically involves a represented perspective—as we'll see below in discussing the Dependency Thesis, we can think of this perspective as the *experiential* perspective represented in memory. What we need to ask here is: Must this represented perspective be a *spatial* perspective which the recalled contents are represented as related to, and so as spatial in turn? Doubts on this score are most easily appreciated by consideration of putative counterexamples.

Discussions of episodic memory are often visuocentric (witness the use of the term 'viewpoint' as a synonym for 'perspective' in the second passage quoted above). But we can of course enjoy non-visual memories. For instance, you might smell a powerful aroma and then later recall having done so. Having once smelled ripe durian fruit, you might never forget smelling its pungent odour of 'turpentine and onions, garnished with a gym sock' (Sterling 2003: 102). Likewise, you might hear a distinctive tune and long recall hearing it thereafter. Perhaps you can recall once hearing a thrush nightingale sing.

Such memories may involve rich spatial scene construction—perhaps you recall yourself smelling the fruit in the midst of a bustling Bangkok market, or hearing the bird in the thick of a Romanian forest. But it is far from obvious they must. Simply to recall that particular experience of the pungent odour, or that particular hearing of the bird's resonant high-pitched *piuu*, *piuu*, does not obviously require representing oneself as occupying any particular spatial location, nor spatially relating any features to such a location. This is not because olfaction or audition in general

[12] Two further puzzles arise concerning the argument. The first concerns the assumption that the assumed viewpoint will inevitably be *the self*. The second concerns the apparently differing attitudes to space and time. For whilst the assumed viewpoint is initially introduced as 'a locus in space and time', it is only the spatial location of the viewpoint which is subsequently drawn on to distinguish event from semantic memory.

lack spatial content.[13] It is because olfactory and auditory memories may sometimes fail to specify any such content. Mohan Matthen contends: 'Auditory memory is hardly ever spatial: usually when you recall a tune, you don't hear it as coming from anywhere' (2010: 14). I make no claim about frequency, which may well vary by individual, but Matthen is plausibly right about the simple possibility of such memories.

Examples of non-spatial memories in non-visual senses may not convince everyone. One might suspect, for instance, that they simply involve extremely generic spatial content (the tune coming from 'somewhere' or 'nearabouts', etc.). Or one might be concerned that when fully deprived of spatial content, such memories become general event memories as opposed to episodic memories proper. Fortunately, we can set aside these concerns, since there is a much clearer case where spatial scene construction can be absent from episodic memory, namely memories of our mental actions, and specifically our thoughts.

As Michael Hasselmo describes at the start of his wonderful book, *How We Remember*, what he calls 'episodic trajectories are not necessarily limited to the dimensions of physical time and space … I also remember my thoughts' (2012: 7). In remembering, in Browning's phrase, being 'stung by the splendour of a sudden thought', one need not remember any spatial content. You can simply remember having the thought. It makes perfect sense to say, 'I have no memory at all of where or when it was, of what happened before or after etc., but I vividly remember thinking …'. To allow for the possibility of such memories, we must deny that all episodic memories constitutively have spatial content. In consequence, Rubin and Umanath's claim that scene content distinguishes episodic from semantic memory must be rejected.[14]

I do, however, think that an important idea in the vicinity of Rubin and Umanath's is correct. This is the idea that memory is a *second-level* phenomenon, in that it involves the representation of a prior experience of yours.[15] Thus, at the first level we have some form of experience: seeing a scene, hearing a tune, thinking a thought. Then, at the second level we represent not merely the scene, tune or thought, but the seeing, hearing or thinking of the thought.[16] This idea is sometimes called the Dependency Thesis. According to it, to paraphrase Martin (2001), to episodically

[13] Reid notoriously makes this claim about olfaction, suggesting: 'It is evidently ridiculous, to ascribe to [a smell or odour] figure, colour, extension, or any other quality of bodies. [One] cannot give it a place, any more than [one] can give a place to melancholy or joy' (1764/1997, I.i.2). For criticism of this tradition, see Richardson (2013) who argues that we do perceive odours as external to our bodies, specifically as in the vicinity of our noses and as brought into our noses by sniffing. For discussion of the spatial content of audition, see Nudds 2009.

[14] A possible line of objection is that all memories include some awareness of one's body, and so even in remembering a thought we necessarily remember it in the context of some background of bodily sensation. Specifically, Marcel Kinsbourne argues that a representation of the self is essential to episodic memory and claims that it is bodily awareness which 'puts the stamp of personal experience on the scene' (1995: 218). What is unclear, however, is why we should think of *bodily* awareness as either necessary or sufficient to secure the link between remembered experience and self.

[15] For detailed discussion of two-level accounts of experience more generally, see Martin 2019.

[16] Note that where the first-level phenomenon involves acquaintance, we can think of memory as renewing such acquaintance. However, no commitment is made here to the idea that *all* episodic memory involves renewed acquaintance. Indeed, recognizing that we can episodically remember our thoughts puts pressure on that idea to the extent that one might want to resist the idea that thought involved acquaintance.

remember an event is to remember previously perceptually experiencing that event or being the conscious agent of it.[17]

The Dependency Thesis helps us see what is right and wrong about Rubin and Umanath's argument, according to which episodic memory essentially involves a spatial viewpoint on a scene. What is right is that a distinctive feature of episodic memory is that it embeds a 'viewpoint': the subjective perspective of an experience of the recalled content. But what is wrong is the idea that this viewpoint need necessarily bring with it any spatial content. It will only do so if the first-level viewpoint must necessarily be represented as a spatial viewpoint. But our cases of non-visual memory, and above all memories of thoughts, suggest that this need not be the case. Arguably, the idea that episodic memory is essentially spatial thus reflects a recognition of Dependency combined with an overly narrow focus on certain kinds of visual memory where an essential connection with space is more plausible. In general, however, episodic memory is not constitutively spatial.[18]

Dependency also allows us to see what is right and wrong about the common idea that the phenomenology of memory is one of perceiving again. For instance, Fabrice Teroni suggests that the 'phenomenology distinctive of memory' involves 'it's being for the subject as if she perceived particular events again' (2017: 27). Taken at face value, Teroni's claim here is false. Episodic recall is not a form of hallucination. Nonetheless, his claim rightly reflects the internal connection between the phenomenology of episodic recall and perception, one captured by the dependency of the former on the latter. What is right from the perspective of the Dependency Thesis is that there is an internal connection between the phenomenology of perception and the phenomenology of memory—to remember is to remember perceiving (or acting). What is wrong is to suggest that they have the very same structure. They do not. Perception involves direct acquaintance with the environment. Memory involves the representation of such acquaintance and thereby the experienced environment.

[17] As formulated, Dependency claims that to remember is to remember *one's own* perceptual experience of an event etc. This raises the question of what to say about so-called *observer* as opposed to *field* memories: memories in which we recall a past episode from a perspective distinct from that which we experienced it, perhaps even looking at ourselves identified as such (see Nigro and Neisser 1983). One option here is to think of the perspectives taken-up in these memories as genuinely our own viewpoints on the grounds that our perceptions contain both egocentric and allocentric spatial information (for a rich discussion, see McCarroll 2018). A more straightforward reply is to think of such memories as immediate cousins of episodic memories, though not strictly episodic memories proper. One reason to insist on a close connection is that, like field memories, they arguably secure a link back to our own experience insofar as the content of any experience represented in an observer memory must be properly inherited from an earlier experience of our own. This allows for transformations to new perspectives in observer memories, whilst preserving what Martin calls the *Previous Awareness Condition*, the idea that 'one can remember an event only where one previously witnessed it or was the conscious agent of it' (2001: 261). Note further that the 'fact that certain *elements* of an experience depend on some present reconstruction [e.g., the transformation to a novel perspective] does not imply that the whole experience could not possibly count as a memory at all' (Debus 2007: 197, and her discussion in §5 more generally).

[18] Might some extended or abstract notion of space (as in: pitch or color space) offer a way of recovering a version of Rubin and Umanath's claim that episodic memory is constitutively spatial? The existence of memories for thoughts again makes me sceptical.

Is Episodic Memory Constitutively Temporal?

Even if episodic memory is not constitutively spatial, one might think it far more obvious that it has a constitutive connection to time. Here, however, it is important to distinguish a number of quite different connections which one might posit between episodic memory and time. With these distinctions in mind, I now argue, contrary to various authors, that episodic memory has only a minimal constitutive connection to time. I consider three putative connections. First, the idea that episodic memories necessarily represent *when* their events occurred. Second, the idea that episodic memories necessarily have *tensed* contents, representing their contents *as past*. Third, the idea that episodic memories necessarily have *temporal contents*, for instance, representing succession, duration and change.

Remembering When

A first way in which episodic memory might be thought constitutively temporal is if it necessarily represented *when* its events occurred, either in terms of absolute location or distance from the present moment. However, despite continuing claims to the contrary[19], it is now a familiar point that it is 'possible, though rare, for one to re-experience an event memorially but have no robust knowledge of how long ago it happened' (Russell and Hanna 2012: 32).[20] Indeed, a case can be made for the stronger view on which our knowledge of *when* a recalled event occurred is largely, if not exclusively, a matter of non-temporal features and background knowledge.[21] Here we might compare memory to a photograph collection in which the photos lack labels, date stamps, or chronological organization, leaving one to figure out from various clues (e.g., the type and fading of the photograph, background knowledge about the fashions and hairstyles people are sporting, etc.) when each photo was taken. I return to the question of the association between episodic memory and temporal organization in the final section. For now, I simply accept the minimal point that some episodic memories do not encode substantial information about when the remembered event occurred.

[19] For instance, Smith and Mizumori: 'episodic memories, by definition, include information about the time and place where the episode occurred' (2006: 716). Cf. Clayton and Dickinson (1998) on episodic-like memory as what–where–when memory.

[20] Boyle puts the point more strongly, holding that there are 'a great many cases in which we have episodic memories for events we are unable to temporally locate' (2021: 2515).

[21] See, e.g., Rubin and Umanath: 'The time of an event need not be known, either in absolute terms or relative to other events; deciding on when an event occurred is a separate set of processes from recalling other properties of the event' (2015: 2). Also, Friedman (1993): 'time information is stored … in our more general body of knowledge about time patterns' (see also his 2004; Brewer 1986, 1996; Rubin & Baddeley 1989; St. Jacques et al. 2008). For critical discussion, see Brown and Chater 2001: 102ff.

Remembering as Past

Even if memories do not always encode *when* a remembered event occurred, one might nonetheless think that memories at least represent events *as past*, and perhaps locate when they occurred in that very minimal sense (i.e., before *now*, the time of recall). Indeed, the claim that whereas perception represents events *as present*, episodic memory represents events *as past* is ubiquitous.[22] Here, however, and drawing on Martin's (2001) discussion of the Dependency Thesis, I want to suggest a more minimal picture which I suggest captures the intuitive data and yet denies that episodic memories themselves have past tensed content. At the very least, I take this more minimal picture to show that the claim that memories represent events in a tensed fashion (i.e., as past) is not simply something that can be taken for granted without argument.

According to the Dependency Thesis, episodic memory represents *particular* experiences. Now, to the extent that we are in position to know what experiences we are undergoing when we are undergoing them, when we remember a particular experience, we are in a position to know (a) that we are not *then* undergoing it, and (b) that it is nonetheless the representation of a *particular* experience.[23] Thus, we can know that our memory is not of a present experience, but of some other particular experience. This leaves open two possibilities: either the recalled experience occurred in the past or it will occur in the future. However, since the represented experience is a *particular* experience, we can be sure, given the basic causal structure of the universe we live in, that it must be located in our pasts. For only then could it have left its causal mark on us in such a way that we could now be reacquainted with *that* event in particular.[24] Our universe does not admit of precognition.[25] The upshot is that we can know that episodic memory links us to the past without any need to postulate tensed content.[26] Put another way, the contents of our

[22] A small sampling from philosophical discussions: 'In the memory mode, the content is presented as true with respect to a past perceptual situation, hence the scene represented is felt as past' (Recanati 2007: 141–42); 'Episodic memories … represent themselves as having a certain causal history, namely … as coming from past perceptions of objective facts' (Fernández 2016: 636–37; cf. Searle 1983); 'memory-experience presents itself as about the past' (Matthen 2010: 8), 'Episodic memory is, in and of itself, an experience of an image as in the past' (ibid.: 11); 'We can think of episodic memory as an explicit representation of the past. Its functional value is rooted in its representation of a past experience itself, that is, a representation of the past *as past*' (Droege 2013: 183). Brown and Chater: 'episodic remembering seems to require the ability to represent an event as having happened at a particular time in the past' (2001: 77, citing McCormack and Hoerl 1999). Finally, Russell: 'what we remember … appears as past and not as present' (1912/2001: 26).

[23] In paradigm cases of episodic recollection, we know that we are recalling a particular episode as opposed to imaging an event type. But it is by no means obvious that this is always the case. Certainly, sometimes we may confuse genuine memory for mere imagination, and so fail to appreciate the particularity of the representation. However, such cases alone do not suffice to show that we are not always *in a position to know* that we are recalling not imagining.

[24] This point also rules out our memory being of a merely possible particular event.

[25] A link to our *own* past, as opposed to someone else's, may also be secured in like manner. On this view, episodic memories would not represent past experiences as *ours*. Instead, our recognition that such memories must be from our pasts would be grounded in an implicit appreciation of the fact that a link to a particular past experience could not be secured by a chain or trace which took us outside of our own mind. Thus, just as our universe contains no precognition, it also admits no telepathic transmission.

[26] For further discussion, see Martin 2001. Of course, such thoughts are not articulated by most people in their ordinary thinking about memory; rather they make explicit what is implicit in and justifies our ordinary thinking. Note that generic event memories (e.g., reading bedtime stories to one's child when young etc.) can still be

memories in and of themselves do not determine that they are not 'memories' of future particular experiences. This is ruled out instead by very general background causal facts about our world.

Another analogy with photography may be helpful here (although as with all analogies it has its limits). Imagine taking an old photograph back to the place where it was taken, perhaps with the aim of taking a second photograph as in Figure 1.2. Holding up the original photograph as the photographer has done here, they may be in a position to know that it depicts events which occurred in the past. But to explain this, we need not attribute tensed contents to photographs. Instead, the photographer can simply reflect on the contrast between what he can currently see and what is depicted in the photograph, a contrast which suffices to show that the photograph does not depict things as they are now. He can then further reflect that the way our world works precludes taking photographs of the future, for photography is a causal and so forwards-only process, which thus can only record particular past occurrences and scenes. In this way, our photographer can know that his photograph depicts the past. There is only re-photography, not pre-photography.[27]

Figure 1.2 'Looking into the past—take 2'. © Nomad Tales. Attribution-ShareAlike 2.0 Generic (CC BY-SA 2.0). Example of rephotography. Original photograph of flooding on the High Street, Maitland, NSW taken in 1955, second photograph of same high street, taken 2010.

The view here is not intended to impose substantive intellectual requirements on episodic memory itself, and so again has no implications for the distribution of episodic memory across humans and non-human animals. It does though imply that appreciating that our episodic memories relate to our pasts (as opposed to our futures or to merely possible events) is a cognitive achievement which goes beyond simply having such memories. Just as the capture, use and collection of photographs

understood as located in our pasts, insofar as they are not purely general, but rather genericizations of particular events.

[27] Note here the contrast with paintings which can depict future events, but cannot, independent of the intentions of their painter, depict particular events (see here Martin 2001: 276). As with remembering, note that it is a non-trivial achievement to appreciate that a print is indeed a photograph (as opposed, e.g., to a print of an entirely AI-generated image) and so depicts a particular.

is possible without any general understanding as to why photographs always depict the past, so too a creature need not have any capacity to make explicit the reasoning offered here in order to form, use and store episodic memories. One might object here that episodic memories have a distinctive phenomenology, an immediate 'felt-as-past' character, as it were. The view here denies this. What is critical to episodic memory's connection to the past is the particularity of its content and not its being past tensed.

Remembering and Internal Temporal Content

Thus far, our discussion has been largely negative. However, I now want to explore a more positive connection between episodic memory and time, a connection which again relates to the Dependency Thesis discussed above.

Often the objects of memory are extended episodes: a live performance, a wedding speech, a dramatic fall. In remembering them, we often represent their temporal features such as succession, duration or change. However, we should allow for the possibility of snapshot memories.[28] Thus, in general, we should not insist that memories must have such temporal features amongst their contents. On the other hand, by Dependency, in the case of perceptual memories, when we remember a past event, such as a performance, we remember previously perceptually experiencing that event. This means that the contents of such perceptual memories do always involve a minimal temporal aspect. This is because we not only remember the performance which occurred and our hearing and seeing of it, but these as temporally related to one another.

Elsewhere, I have defended a particular view about the relationship between the timing of perceptual experience and the timing of its objects.[29] According to this view, temporal features of experience are *inherited* from their objects. More precisely: for any temporal property apparently presented in perceptual experience, experience itself has that same temporal property. In particular, the apparent temporal location and duration of an event determines the temporal location and duration of our experience of said event. Inheritance articulates the idea that part of what it is like to perceive the world from a perceiver's own perspective is to be temporally yoked to the world. Our stream of consciousness is manifestly concurrently shaped by the events we consciously perceive.

By Dependency, perceptual memories involve representing being so temporally yoked to the world. We thereby represent our stream of consciousness being determined in its temporal location and duration (if any) by the recalled event. To that extent, episodic memory involves the representation of a temporal structure: an

[28] Compare Tulving, who describes his General Abstract Processing System framework as offering a '"snapshot view" of episodic memory [focusing] on the conditions that bring about a slice of experience frozen in time which we identify as "remembering"', further proposing that episodic memory 'produces many snapshots whose orderly succession can create the mnemonic illusion of the flow of past time' (1984: 231).

[29] See Phillips 2009, 2010, 2014a, reviewed in Phillips 2014b. For closely related ideas, see Soteriou 2010 and 2011.

experience and its object, occurring (and perhaps unfolding) together in time. Recall Rubin and Umanath's argument that episodic memory essentially involves a spatial viewpoint on a scene. I rejected this contention partly on the ground that we can have non-spatial (e.g., auditory and olfactory) memories. However, we can now see that there is a sense in which all perceptual memories do essentially involve a temporal 'viewpoint': the subjective temporal perspective of our experience of the recalled episode. There is then a minimal respect in which episodic memory represents when an event occurs: it represents the event perceived as occurring *contemporaneously* with our perceiving of it (and indeed as determining the temporal location of that perceiving).[30]

These claims relate specifically to perceptual memory. As against Rubin and Umanath's claim that episodic memory is constitutively spatial, I objected that we can also remember our thoughts. The claims just made do not straightforwardly extend to the case of recalled thought. For no analogue of inheritance applies to thinking; the temporal locations and durations of our thoughts are not inherited from their objects—you can ruminate for hours on a momentary incident, or for just a moment about an epoch. On the other hand, in the case of conscious thought it is far from obvious that we can genuinely make sense of snapshot memories. To remember thinking, in a genuinely episodic manner (as opposed to simply that one had a certain thought), it might seem necessary to involve recalling some (at least briefly) extended episode.[31] If this is right, then a more general claim about episodic memory can still be made, namely that episodic memory constitutively has some internal temporal content: minimally, duration in the case of recalled thought, and simultaneity in the case of recalled perception.

Space, Time, and the Organization and Encoding of Episodic Memory

In this final section, I briefly consider whether what is known about the encoding and organization of *human and mammalian* episodic memory motivates a stronger link between memory and spatiotemporal representation than the minimal connection so far acknowledged.

In approaching this question, we must centre a feature of memory hitherto neglected, namely storage and retrieval. Memory is not simply a depository where event-representations are stored without any particular structure or system: a giant

[30] The discussion here is indebted to Soteriou's (2018) rich defence of the idea that episodic memory constitutes a form of 'mental time travel'. Soteriou argues that episodic memory provides a 'way of representing entities as temporally present' (2018: 308) which is distinct from past tensed thought. Specifically, he argues that in representing a perceptual perspective on a past event (as Dependency holds one does in episodically remembering), one represents a temporal perspective whose temporal location is determined by the temporal location of the past event itself. As a result, one represents a perspective distinct from one's actual present perspective to which the past event is present. I take the picture defended here to be close to Soteriou's. However, it is neutral on whether there is strictly any tensed representation in perception or memory (i.e., representation of events as temporally present).

[31] For an argument that conscious thinking necessarily involves mental events with duration, see Soteriou 2009.

pile of photographs tossed into a mental junk drawer. Rather, as Aronowitz powerfully argues, a 'core function of any memory system is to support accurate and relevant retrieval' (2019: 483). And if memories are to be retrievable in a timely and contextually appropriate manner, they must be *organized*.

As Aronowitz (2019) discusses, there are important questions here as to how such retrieval efficiencies are achieved. For example, do our memory systems index an unorganized store, or organize memories into simplified models, or both? But in either case, it will be necessary to use general, abstract 'contextual' features in indexing or modelling. Moreover, just as people typically organize their photographs chronologically, it is natural to ask whether space and time provide special, perhaps even naturally necessary, organizational dimensions in relation to episodic memory. If so, whilst particular episodic memories might lack spatiotemporal content, in general, the association of episodic memories with information about spatiotemporal context would be critical to the proper functioning of the episodic memory system. To take up our analogy with photographs once more, memory may be like a chronologically organized photograph collection, in which, occasional stray photographs aside, all collected photographs are associated at least with a relative temporal position.

As discussed by Brown and Chater (2001, drawing on Anderson 1990), there are clear adaptive reasons to suppose that memories will be organized spatiotemporally. For suppose that retrieval is costly in terms of time and/or limited resources, then it will be important to ensure that what is retrieved is maximally relevant. Moreover, space and time are both externally and internally linked to relevance. In terms of external linkages, and all else equal, the relevance of information declines monotonically with distance from our current location. Arriving in Rome, we want to recall how the Romans do things, not the Greeks, still less the Egyptians. Similarly, for duration. In today's Rome, information about current customs and conditions is likely much more relevant than information about those during the 1960s, let alone the height of the Republic.[32] In terms of internal linkages, and again all else equal, it is plausible that having accessed a memory concerning some location or time, it is more likely that information concerning nearby times and places will be relevant (cf. Aronowitz 2018 on temporally ordered retrieval, as in the familiar—but effective—routine of trying to find a lost item by asking when you last saw it).[33]

These considerations only take us so far. They show that there is adaptive value in organizing memories spatiotemporally. But, as Brown and Chater concede (2001: 102), they do not show that memories will exclusively exhibit spatiotemporal organization. Indeed, similar adaptive arguments could be made in support of organization along many other dimensions. Suppose I am trying to recall a tune. Since tunes tend to continue with notes nearby in pitch, having recalled a note, then all else

[32] More precisely, and again all else equal, relevance declines monotonically with time since last *retrieval* (Anderson and Schooler 1991).

[33] Of course, all else is often not equal, and context is crucial. Thus, information about the Roman Republic might be a better guide to the politics of Washington, D.C. today than information about the east coast of America even a thousand years after the fall of Rome. (Thanks to Simon Brown for the example.)

equal, information concerning continuations close by in pitch will be more relevant. Or, to anticipate an example below, suppose I encounter (or recall) someone who occupies a particular social standing with respect to me, a powerful friend, or a helpless stranger. All else equal, it will be most relevant to recall information concerning encounters with others located similarly along such social dimensions of power or affiliation. For I will want to recall how such people tend to act, what they tend to like or need etc. Now, of course, unlike a physical photograph album, nothing prevents memories being organized along multiple dimensions, either in the sense of being searchable using a range of different dimensions, or in the sense of exhibiting a pattern of activation dependencies along the lines discussed above. But once this very general point is recognized, it is unclear what would justify the assumption that time is 'always a factor' as Brown and Chater hold (2001: 102). Instead, the possibility arises not only that certain memories may be organized in non-spatiotemporal ways but that some memories may not be organized spatiotemporally at all.

One important reason to think that space and/or time are always a factor is that episodic memory systems have arguably evolved out of simpler systems apparently dedicated to spatial navigation (O'Keefe and Nadel 1978; Buszáki 2005) and time-dependent foraging (Gallistel 1990; Brown and Vousden 1998). Given this phylogeny, we might expect systems like ours to exhibit spatiotemporal organization, even if in principle other organizations are possible. Space and time might then be the basic or default mode of organization of human or mammalian memory.

According to a highly influential proposal supported by a wealth of studies on amnesic patients, hippocampal structures play a critical role in the coding of human episodic memories (e.g., Vargha-Khadem et al. 1997; Tulving and Markowitsch 1998; Aggleton and Brown 1999). Yet, as we know from rodent studies, hippocampus also plays a critical role in spatial navigation, being the site of so-called 'place cells': cells which fire at high frequency when an animal occupies a particular region of space and are thought to provide the basis for a rodent's 'cognitive map' of their spatial environment—a viewpoint independent representation (or better: overlapping set of representations) of the relationships between features and locations in the animal's environment (O'Keefe and Dostrovsky 1971; O'Keefe and Nadel 1978; Mizumori 2008).[34] So-called 'time cells' have also now been discovered in hippocampus and entorhinal cortex (Pastalkova et al. 2008; MacDonald et al. 2011; Kraus 2015; Salz 2016; Umbach 2020). These are cells which fire at specific time points in a learned sequence even when the animal is stationary.

Together these findings raise the vexed question of how precisely such spatial and temporal mapping functions in rodents relate to human episodic memory. As Kathryn Jeffery argues: 'If one assumes that there is a functional homology

[34] Several other types of spatially responsive cells are now recognized in hippocampus and entorhinal cortex, e.g., head-direction, boundary, and grid cells. For review, see Hartley et al. 2014. Note that as discussed shortly below it should not be assumed that any of these cells *only* encode information about space. Nor should it be assumed that spatial navigational processing is in any sense limited to such brain regions. There is clear evidence that this is not the case, as discussed by Ekstrom and Ranganath (2018: 680–81).

between rodents and humans with regard to the anatomy and mechanisms of episodic memory, then spatial and episodic functions need somehow to be unified' (2008: 69). Moreover, evidence of a broadly homologous basis for *spatial navigation* in humans (Ekstrom et al. 2003) raises the question directly of how episodic memory and navigational functions relate in humans given their apparently shared neuroanatomical basis.

There is now a great deal of evidence that non-spatiotemporal information is encoded by place and other hippocampal and entorhinal cortex cells. For instance, hippocampal cells have been found to respond strongly to many task-relevant stimuli such as odours, colours and visual-tactile cues (Eichenbaum et al. 1987; Igarashi et al. 2014; Anderson and Jeffery 2003; Young et al. 1994; reviewed in Eichenbaum et al. 1999). One way to understand this evidence is that hippocampal circuits[35] provide the basis of a spatiotemporal cognitive map which 'is richly embellished by nonspatial information to form a representation of [spatiotemporal] context, used to (among other things) organize memories' (Jeffery 2008: 65).[36] According to this proposal, episodic memories are organized by being '"attached" to a map' (ibid.: 69), where this map is fundamentally spatiotemporal, even if other features are encoded as part of a richer spatiotemporal context. On this view, then, whilst individual episodic memories themselves need not have spatiotemporal contents, a spatiotemporal framework forms the fundamental basis of episodic memory organization, subserving its critical retrieval functions.

However, an alternative account denies that hippocampal encoding is inextricably spatiotemporal. On this alternative, hippocampus maps relations quite generally, encoding relationships between myriad encountered sensory features across all relevant continuous dimensions. It is conceded that in everyday life, and perhaps especially in traditional experimental tasks (e.g., those involving maze exploration), that spatial information is typically such a reliable identifier of context 'that its inclusion in context representations is largely automatic' (Smith and Mizumori 2006: 721), leading to the appearance that space is privileged. However, strictly speaking, 'spatial information is but one of the many kinds of information that serves the general context processing function of the hippocampus' (ibid.: 727). Thus, memory organization fundamentally involves an 'interleaving of events and episodes into relational networks' with space and time just being examples amongst equals of potential organizational relations (Eichenbaum and Cohen 2014: 764). From this perspective, 'place cells' (etc.) are misnamed since they encode many other kinds of contextual information depending on the task.

What considerations might be adduced in favour of this latter perspective? Aronov et al. (2017) trained stationary rats to release a joystick when a sound reached a fixed

[35] Henceforth, I omit reference to other connected regions such as entorhinal and parahippocampal cortex.

[36] Jeffery focuses only on spatial context, whereas for reasons given in the main text, I broaden this to spatiotemporal context. Note that, though I do not explore it here, it is possible to resist the move beyond space to time in the manner discussed below with respect to non-spatiotemporal dimensions.

frequency range, decoupling pitch from temporal contingencies by randomly varying the speed of the frequency increase. Recordings of hippocampus and medial entorhinal cortex showed cells forming 'frequency fields', analogous to the 'place fields' famously associated with spatial navigation.[37] Aronov et al. take these results to show that 'spatial representation is just one example of a more general mechanism for encoding continuous, task-relevant variables' (2017: 7). Even more abstractly, Tavares et al. (2015; highlighted by Eichenbaum 2015) had human participants play a role-playing game in which they had to move to a new town and find work and housing. In the game, the characters with which participants interacted all occupied varying positions in an egocentrically defined 'social space' with axes corresponding to dimensions of power and affiliation (e.g., new boss, old friend). Tavares et al. found that hippocampal activity correlated with vector angle in this space, suggesting that hippocampal networks can map even highly abstract, culturally constructed spaces. Again, the authors construe this as evidence that neither space nor time provide a privileged organizational scheme for episodic memory; any abstract structure will do.

These considerations are not decisive. Evidence from such studies is consistent with the view that whilst place cells 'might be interested in more than just space … they are primarily interested in space' (Jeffery 2008: 69). More generally, as Ekstrom and Ranganath argue, extant evidence does not rule out the view that space and time provide the 'primary scaffold for defining contexts, and for organizing incoming information within a context representation' with other dimensions being 'incorporated into the scaffold' as relevant (2018: 685). For instance, concerning Aronov et al.'s task, Ekstrom and Ranganath suggest that when the rats first entered the testing chamber, their hippocampi would have encoded information about the spatial and temporal structure of the chamber and task. Only later, after learning the task-specific pitch contingencies, would this dimension of their experience have been added to their contextual representation (presumably without abolishing ongoing spatiotemporal coding). Another possibility is that information about non-spatial dimensions (frequency, social affiliation) becomes automatically associated with spatial information, with the result that spatial representations are activated in these tasks.[38]

These issues are not yet settled and much remains unclear about the precise role and cognitive function of hippocampus (as well as other areas such as parahippocampal and entorhinal cortex). This said, the success of recent work (e.g., Whittington et al. 2020; Benna and Fusi 2021; Momennejad 2020) in predicting specific patterns of spatial and temporal neural responses simply by modelling the system as aimed at abstracting, generalizable structure from sensory episodes, would most

[37] Place fields are regions of space in which corresponding 'place cells' fire at high frequency. Frequency fields are thus pitch intervals in which corresponding cells fire at a high rate.

[38] Consider here too Walker et al. (2010) who provide evidence that cross-modal associations (such as between pitch and height) are present even in 3- to 4-month-old infants and so arguably reflect an innate aspect of perception. (Thanks here to E. J. Green.)

naturally appear to support a view on which spatiotemporal representations are emergent features of a system which does not accord them any special privilege. On this more parsimonious view of the role of hippocampus as an extractor of abstract, generalizable structure, be it spatial or non-spatial, whilst of course we can expect our memories to exhibit spatiotemporal organization, the sense, if any, in which space and time are basic or default organizational dimensions of episodic encoding will simply be that these are—as a matter of contingent fact—primary organizing dimensions of our perceptual experience. We are spatiotemporal rememberers because we live in and perceive a spatiotemporal world.

Conclusion

I have considered four issues: the nature of episodic memory, whether it is constitutively spatial, whether it is constitutively temporal, and whether space and time provide fundamental principles for organizing episodic memory. In the main, my answers have been negative. Episodic memory does not appear to be constitutively spatial, and its constitutive connection with time is far more minimal than most philosophers believe. Furthermore, whilst space and time do play a critical role in the organization of episodic memory, this role is neither obviously unique nor fundamental. On the other hand, I have defended a modest constitutive connection between episodic memory and time related to the temporal structure of its contents.[39]

References

Aggleton, J.P., and Brown, M.W. (1999). Episodic memory, amnesia, and the hippocampal-anterior thalamic axis. *Behavioral and Brain Sciences*, 22, 425–44.

Anderson, J.R. (1990). *The Adaptive Character of Thought*. Hillsdale, NJ: Erlbaum.

Anderson, J.R., and Schooler, L.J. (1991). Reflections of the environment in memory. *Psychological Science*, 2, 396–408.

Anderson, M.I., and Jeffery, K.J. (2003). Heterogeneous modulation of place cell firing by changes in context. *Journal of Neuroscience*, 23, 8827–35.

Andonovski, N. (2020). Singularism about episodic memory. *Review of Philosophy and Psychology*, 11(2), 335–65.

Aronov, D., Nevers, R., and Tank, D. (2017). Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit. *Nature*, 543, 719–22.

Aronowitz, S. (2019). Memory is a modeling system. *Mind & Language*, 34(4), 483–502.

[39] Thanks to all the participants at the 'Memory, Space, and Time' conference for their comments and questions, and especially to Sara Aronowitz and Lynn Nadel for organizing such an excellent occasion. Special thanks to Ali Boyle for a very helpful conversation following my talk, and to Simon Brown, E. J. Green, and Matthew Soteriou for hugely helpful written comments on an earlier draft which led to substantial improvements to the paper.

Aronowitz, S. (forthcoming). Semanticization challenges the episodic-semantic distinction. *British Journal for the Philosophy of Science*.

Baddeley, A. (2002). The concept of episodic memory. In A. Baddeley, J. Aggleton, and C. Martin (eds.) *Episodic Memory: New Directions in Research* (pp. 1–10). Oxford: Oxford University Press.

Bartlett, F.C. (1932). *Remembering: A Study in Experimental and Social Psychology*. Cambridge: Cambridge University Press.

Benna, M.K., and Fusi, S. (2021). Place cells may simply be memory cells: Memory compression leads to spatial tuning and history dependence. *PNAS*, 118(5), 1–12.

Block, N. (1995). On a Confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18(2), 227–47.

Block, N. (2023). *The Border Between Seeing and Thinking*. Oxford: Oxford University Press.

Boyle, A. (2021). Remembering events and representing time. *Synthese*, 199, 2505–24.

Brewer, W.F. (1986). What is autobiographical memory? In D.C. Rubin (ed.), *Autobiographical Memory* (pp. 25–49). Cambridge: Cambridge University Press.

Brewer, W.F. (1996). What is recollective memory? In D.C. Rubin (ed.), *Remembering Our Past: Studies in Autobiographical Memory* (pp. 19–66). Cambridge: Cambridge University Press.

Brown, G.D.A., and Chater, N. (2001). The chronological organization of memory: Common psychological foundations for remembering and timing. In C. Hoerl and T. McCormack (eds.), *Time and Memory: Issues in Philosophy and Psychology* (pp. 77–110). Oxford: Oxford University Press.

Brown, G.D.A., and Vousden, J.I. (1998). Adaptive analysis of sequential behavior: oscillators as rational mechanisms. In M. Oaksford and N. Chater (eds.), *Rational Models of Cognition* (pp. 165–93). Oxford: Oxford University Press.

Brown, S.A.B. (manuscript). What if anything is semantic memory like in non-human animals?

Burge, T. (2010). *Origins of Objectivity*. Oxford: Oxford University Press.

Burge, T. (2022). *Perception: First Form of Mind*. Oxford: Oxford University Press.

Buszáki, G., and Moser, E. I. (2013). Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nature Neuroscience*, 16(2), 130–38.

Buzsáki, G. (2005). Theta rhythm of navigation: link between path integration and landmark navigation, episodic and semantic memory. *Hippocampus*, 15(7), 827–40.

Campbell, J. (2002). *Reference and Consciousness*. Oxford: Oxford University Press.

Campbell, J. (2011). Review: Origins of Objectivity by Tyler Burge. *Journal of Philosophy*, 108(5), 269–85.

Carey, S. (2009). *The Origin of Concepts*. Oxford: Oxford University Press.

Clayton, N.S., and Dickinson, A. (1998). Episodic-like memory during cache recovery by scrub jays. *Nature*, 395(6699), 272–74.

Copenhaver, R. (2006). Thomas Reid's Theory of Memory. *History of Philosophy Quarterly*, 23(2), 171–89.

De Brigard, F. (2014). Is memory for remembering? Recollection as a form of episodic hypothetical thinking. *Synthese*, 191, 1–31.

Debus, D. (2007). Perspectives on the past: A study of the spatial perspectival characteristics of recollective memories. *Mind & Language*, 22(2), 173–206.

Droege, P. (2013). Memory and consciousness. *Philosophia Scientiæ*, 17(2), 171–93.

Eichenbaum, H. (2015). The hippocampus as a cognitive map … of social space. *Neuron*, 87(1), 9–11.

Eichenbaum, H., and Cohen, N.J. (2014). Can we reconcile the declarative memory and spatial navigation views on hippocampal function? *Neuron*, 83, 764–70.

Eichenbaum, H., Dudchenko, P., Wood, E., Shapiro, M., and Tanila, H. (1999). The hippocampus, memory, and place cells: Is it spatial memory or a memory space? *Neuron*, 23(2), 209–26.

Eichenbaum, H., Kuperstein, M., Fagan, A., and Nagode, J. (1987). Cue-sampling and goal-approach correlates of hippocampal unit activity in rats performing an odor-discrimination task. *Journal of Neuroscience*, 7, 716–32.

Ekstrom A.D., Kahana M.J., Caplan J.B., Fields T.A., Isham E.A., Newman E.L., and Fried, I. (2003). Cellular networks underlying human spatial navigation. *Nature*, 425, 184–88.

Ekstrom A.D., and Ranganath, C. (2018). Space, time, and episodic memory: The hippocampus is all over the cognitive map. *Hippocampus*, 28(9), 680–87.

Fernández, J. (2016). Epistemic generation in memory. *Philosophy and Phenomenological Research*, 92, 620–44.

Firestone, C., and Scholl, B. (2016). Cognition does not affect perception: Evaluating the evidence for 'top-down' effects. *Behavioral and Brain Sciences*, 39, 1–19.

French, C., and Phillips, I. (2023). Naïve realism, the slightest philosophy, and the slightest science. In B. McLaughlin and J. Cohen (eds.), *Contemporary Debates in the Philosophy of Mind* (pp. 363–83). Hoboken, NJ: John Wiley & Sons Ltd.

Friedman, W.J. (1993). Memory for the time of past events. *Psychological Bulletin*, 113(1), 44–66.

Friedman, W.J. (2004). Time in autobiographical memory. *Social Cognition*, 22(5), 605–21.

Gallistel, C.R. (1990). *The Organization of Learning*. Cambridge, MA: MIT Press.

Gibson J.J. (1979). *The Ecological Approach to Visual Perception*. Boston, MA: Houghton, Mifflin and Company.

Gregory, R.L. (1966). *Eye and Brain: The Psychology of Seeing*. New York: McGraw-Hill.

Hampton, R.R., and Schwartz, B.L. (2004). Episodic memory in nonhumans: What, and where, is when? *Current Opinion in Neurobiology*, 14, 192–97.

Hartley, T., Lever, C., Burgess, N., and O'Keefe, J. (2014). Space in the brain: How the hippocampal formation supports spatial cognition. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 369(1635), 1–18.

Hasselmo, M.E. (2012). How We Remember: Brain Mechanisms of Episodic Memory. Cambridge, MA: MIT Press.

Igarashi, K.M., Lu, L., Colgin, L.L., Moser, M.B., and Moser, E.I. (2014). Coordination of entorhinal-hippocampal ensemble activity during associative learning. *Nature*, 510(7503), 143–47.

Jeffery, K.J. (2008). The place cells—Cognitive map or memory system? In S.J.Y. Mizumori (ed.), *Hippocampal Place Fields: Relevance to Learning and Memory* (pp. 59–72). Oxford: Oxford University Press.

Kinsbourne, M. (1995). Awareness of one's body. In J.L. Bermúdez, A. Marcel, and N. Eilan (eds.), *The Body and the Self* (pp. 205–23). Cambridge, MA: MIT Press.

Knill, D.C., and Richards, W. (eds.) (1996). *Perception as Bayesian Inference*. Cambridge: Cambridge University Press.

Kraus, B.J., Brandon, M.P., Robinson, R.J., Connerney, M.A., Hasselmo, M.E., and Eichenbaum, H. (2015). During running in place, grid cells integrate elapsed time and distance run. *Neuron*, 88, 578–89.

MacDonald, C.J., Lepage, K.Q., Eden, U.T., and Eichenbaum, H. (2011). Hippocampal 'time cells' bridge the gap in memory for discontiguous events. *Neuron*, 71, 737–49.

Mandelbaum, E. (2018). Seeing and conceptualizing: Modularity and the shallow contents of perception. *Philosophy and Phenomenological Research*, 97(2), 267–83.

Marr, D. (1982). *Vision*. San Francisco: W.H. Freeman.

Martin, M.G.F. (2001). Out of the past: Episodic recall as retained acquaintance. In C. Hoerl and T. McCormack (eds.), *Time and Memory: Issues in Philosophy and Psychology* (pp. 257–84). Oxford: Oxford University Press.

Martin, M.G.F. (2015). Old acquaintance: Russell, memory and problems with acquaintance. *Analytic Philosophy*, 56(1), 1–44.

Martin, M.G.F. (2019). Betwixt feeling and thinking: Two-level accounts of experience. In J. Knowles and T. Raleigh (eds.), *Acquaintance: New Essays* (pp. 95–126). Oxford: Oxford University Press.

Matthen, M. (2010). Is memory preservation? *Philosophical Studies*, 148(1), 3–14.

McCarroll, C.J. (2018). *Remembering from the Outside: Personal Memory and the Perspectival Mind*. Oxford: Oxford University Press.

McClelland, J.L., McNaughton, B.L., and O'Reilly, R.C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102(3), 419–57.

McCormack, T., and Hoerl, C. (1999). Memory and temporal perspective: The role of temporal frameworks in memory development. *Developmental Review*, 19, 154–82.

Michaelian, K. (2012). Generative memory. *Philosophical Psychology*, 24, 323–42.

Michaelian, K. (2016). *Mental Time Travel: Episodic Memory and Our Knowledge of the Personal Past*. Cambridge, MA: MIT Press.

Mizumori, S.J.Y. (ed.). (2008). *Hippocampal Place Fields: Relevance to Learning and Memory*. New York: Oxford University Press.

Momennejad, I. (2020). Learning Structures: Predictive Representations, Replay, and Generalization. *Current Opinion in Behavioral Sciences*, 32, 155–66.

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

Neisser, U. (1981). John Dean's memory: A case study. *Cognition*, 9(1), 1–22.

Nigro, G., and Neisser, U. (1983). Point of view in personal memories. *Cognitive Psychology*, 15, 467–82.

Nudds, M. (2009). Sounds and space. In M. Nudds and C. O'Callaghan (eds.), *Sounds and Perception: New Philosophical Essays* (pp. 69–96). Oxford: Oxford University Press.

O'Keefe, J., and Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, 34, 171–75.

O'Keefe, J., and Nadel, L. (1978). *The Hippocampus as a Cognitive Map*. New York: Oxford University Press.

Openshaw, J. (2022). Remembering objects. *Philosophers' Imprint*, 22(11), 1–20.

Pastalkova, E., Itskov, V., Amarasingham, A., and Buzsáki, G. (2008). Internally generated cell assembly sequences in the rat hippocampus. *Science*, 321, 1322–27.

Phillips, I. (2009). *Experience and Time*, PhD Thesis, University College London.

Phillips, I. (2010). Perceiving temporal properties. *European Journal of Philosophy*, 18(2), 176–202.

Phillips, I. (2014a). Experience of and in time. *Philosophy Compass*, 9(2), 131–44.

Phillips, I. (2014b). The temporal structure of experience. In D. Lloyd and V. Arstila (eds.), *Subjective Time: The Philosophy, Psychology, and Neuroscience of Temporality* (pp. 139–58). Cambridge, MA: MIT Press.

Phillips, I. (2018). Unconscious perception reconsidered. *Analytic Philosophy*, 59(4), 471–514.

Quillian, M.R. (1966). Semantic memory. Doctoral dissertation, Carnegie Institute of Technology. Reprinted in part in M. Minsky (ed.), *Semantic Information Processing*. Cambridge, MA: MIT Press.

Quilty-Dunn, J. (2022). Perceptual pluralism. *Noûs*, 54(4), 807–38.

Recanati, F. (2007). *Perspectival Thought: A Plea for (Moderate) Relativism*. Oxford: Oxford University Press.

Reid, T. (1764/1997). *An Inquiry into the Human Mind on the Principles of Common Sense*, D.R. Brookes (ed.), University Park: Pennsylvania State University Press.

Reid, T. (1785/2002). *Essays on the Intellectual Powers of Man*, D.R. Brookes (ed.), University Park: Pennsylvania State University Press.

Richardson, L. (2013). Sniffing and smelling. *Philosophical Studies*, 162, 401–19.

Rock, I. (1983). *The Logic of Perception*. Cambridge, MA: MIT Press.

Rogers, B. (2021). Optic flow: Perceiving and acting in a 3-D world. *i-Perception*, 12(1), 1–25.

Rubin, D.C. (2022). A conceptual space for episodic and semantic memory. *Memory & Cognition*, 50, 464–77.

Rubin, D.C., and Baddeley, A. (1989). Telescoping is not time compression: A model. *Memory & Cognition*, 17, 653–61.

Rubin, D.C., and Umanath, S. (2015). Event memory: A theory of memory for laboratory, autobiographical, and fictional events. *Psychological Review*, 122(1), 1–23.

Russell, B. (1912/2001). *The Problems of Philosophy*. With an introduction by John Skorupski. Oxford: Oxford University Press.

Russell, J., and Hanna, R. (2012). A minimalist approach to the development of episodic memory. *Mind & Language*, 27(1), 29–54.

Salz, D.M., Tiganj, Z., Khasnabish, S., Kohley, A., Sheehan, D., Howard, M.W., and Eichenbaum, H. (2016). Time Cells in Hippocampal Area CA3. *Journal of Neuroscience*, 36(28), 7476–84.

Schacter, D.L., and Addis, D.R. (2007). The cognitive neuroscience of constructive memory: remembering the past and imagining the future. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 362(1481), 773–86.

Searle, J.R. (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.

Seth, A. (2021). *Being You: A New Science of Consciousness*. London: Faber & Faber.

Smith, D.M., and Mizumori, S.J. (2006). Hippocampal place cells, context, and episodic memory. *Hippocampus*, 16(9), 716–29.

Soteriou, M. (2009). Mental agency, conscious thinking, and phenomenal character. In. L. O'Brien and M. Soteriou (eds.), *Mental Actions* (pp. 231–52). Oxford: Oxford University Press.

Soteriou, M. (2010). Perceiving events. *Philosophical Explorations*, 13(3), 223–41.

Soteriou, M. (2011). The perception of absence, space, and time. In J. Roessler, H. Lerman, and N. Eilan (eds.), *Perception, Causation, and Objectivity* (pp. 181–206). Oxford: Oxford University Press.

Soteriou, M. (2018). The past made present: Mental time travel in episodic recollection. In K. Michaelian, D. Debus, and D. Perrin (eds.) *New Directions in the Philosophy of Memory* (pp. 294–312). London: Routledge

Squire, L.R. (1992). Declarative and non-declarative memory: multiple brain systems supporting learning and memory. *Journal of Cognitive Neuroscience*, 4, 232–43.

St. Jacques, P., Rubin, D.C., LaBar, K.S., and Cabeza, R. (2008). The short and long of it: Neural correlates of temporal-order memory for autobiographical events. *Journal of Cognitive Neuroscience*, 20(7), 1327–41.

Sterling, R. (2003). In J. Winokur (ed.), *The Traveling Curmudgeon: Irreverent Notes, Quotes, and Anecdotes on Dismal Destinations, Excess Baggage, the Full Upright Position, and Other Reasons Not to Go There* (p. 102). Seattle, WA: Sasquatch Books.

Tavares, R.M., Mendelsohn, A., Grossman, Y., Williams, C.H., Shapiro, M., Trope, Y., and Schiller, D. (2015). A map for social navigation in the human brain. *Neuron*, 87(1), 231–43.

Teroni, F. (2017). The phenomenology of memory. In S. Bernecker and K. Michaelian (eds.), *The Routledge Handbook of Philosophy of Memory* (pp. 21–33). New York: Routledge.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving and W. Donaldson (eds.), *Organization of Memory* (pp. 381–403). Cambridge, MA: Academic Press.

Tulving, E. (1984). Précis of *Elements of Episodic Memory*. *Behavioral and Brain Sciences*, 7(2), 223–38.

Tulving, E. (2001). Origin of autonoesis in episodic memory. In. H. L. Roediger III and J. M. Nairne (eds.), *The Nature of Remembering* (pp. 17–34). Washington, D.C.: American Psychological Association.

Tulving, E., and Markowitsch, H.J. (1998). Episodic and declarative memory: Role of the hippocampus. *Hippocampus*, 8, 198–204.

Umbach, G., Kantak, P., Jacobs, J., Kahana, M., Pfeiffer, B.E., Sperling, M., and Lega, B. (2020). Time cells in the human hippocampus and entorhinal cortex support episodic memory. *PNAS*, 117(45), 28463–74.

Vargha-Khadem, F., Gadian, D.G., Watkins, K.E., Connelly, A., Van Paesschen, W., and Mishkin, M. (1997). Differential effects of early hippocampal pathology on episodic and semantic memory. *Science*, 277(5324), 376–80.

von Helmholtz, H. (1910/1962). *Treatise on Physiological Optics*. Dover (English translation by J.P.C. Southall from the 3rd German edition of *Handbuch der Physiologischen Optik*. Vos).

Walker, P., Bremner, J.G., Mason, U., Spring, J., Mattock, K., Slater, A., and Johnson, S.P. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science*, 21(1), 21–25.

Wheeler, M.A., Stuss, D.T., and Tulving. E. (1997). Toward a theory of episodic memory: The frontal lobes and autonoetic consciousness. *Psychological Bulletin*, 121, 331–54.

Whittington, J.C.R., Muller, T.H., Mark, S., Chen, G., Barry, C., Burgess, N., and Behrens, T. E. J. (2020). The Tolman-Eichenbaum Machine: Unifying Space and Relational Memory through Generalization in the Hippocampal Formation. *Cell*, 183(5), 1249–63.

Young, B. J., Fox, G. D., and Eichenbaum, H. (1994). Correlates of hippocampal complex-spike cell activity in rats performing a nonspatial radial maze task. *Journal of Neuroscience*, 14(11), 6553–63.

2

Developmental Sequences Constrain Models of the Mind

Nora S. Newcombe and Kim V. Nguyen

Researchers in psychology and neuroscience aim to find analytic categories, analogous to the concept of 'element' in chemistry or 'species' in biology, to facilitate our understanding of cognition, emotion, and behaviour. However, arriving at the right categories in any science is challenging because research is necessarily concept laden, i.e., formulated using the very concepts that may turn out to be questionable (Dubova & Goldstone 2023). Additionally, in cognitive science, we are uncertain about whether we are aiming to characterize the mind and brain as composed of distinct modules or as overlapping constructs instantiated in large neural networks with considerable overlap as well as points of difference. There are various criteria to address these issues, including patterns of cross-species variation, whether individual differences are correlated, whether functions share common neural bases, whether functions show double dissociations, and whether functions show coincident or sequential developmental emergence. However, it is rare to see systematic evaluation using multiple criteria. This chapter considers memory and navigation and argues for using developmental emergence as one tool in defining the relations among constructs. We begin by outlining the crucial conceptual distinctions currently used in studying memory and navigation. We discuss how the proposed constructs (semantic and episodic memory, egocentric and allocentric navigation) may be related to each other, first for within-domain pairs and then across domains.

Defining Memory Systems and Navigation Systems

Memory refers to a variety of mental phenomena. For decades, researchers have distinguished between explicit/declarative and implicit/procedural memory, with explicit memory subdivided into semantic and episodic memory (Squire et al. 1993; Squire 2004; Tulving 1983). Semantic memory refers to general knowledge about the world or fact-based knowledge (e.g., dogs bark), and episodic memory involves storage of life events with their temporal and spatial context (e.g., *Fido barked wildly yesterday, standing at our front door when the new mail carrier was walking up to our house*). Although the boundary is not always clear (Renoult et al. 2019), the basic contrast has been compelling and useful.

Nora S. Newcombe and Kim V. Nguyen, *Developmental Sequences Constrain Models of the Mind*. In: *Space, Time, and Memory*. Edited by: Lynn Nadel and Sara Aronowitz, Oxford University Press.
DOI: 10.1093/oso/9780192882547.003.0002

Navigation is sometimes considered an aspect of 'spatial cognition'. However, that term is misleading; navigation and object-centred spatial manipulation are distinct from each other conceptually, evolutionarily, and neurally (Newcombe 2018). Navigation involves encoding where things are in the wider world (e.g., how to walk from my house to the post office) and planning how to move to goals within that world. Object-centred spatial manipulation involves encoding how things are shaped and how we can transform those internal spatial relations (e.g., folding a letter so it fits an envelope). Navigation depends on two systems, one using body-based information and the other using world-based information (Ekstrom & Isham 2017; Newcombe 2018; Tansan et al. 2022). There are two kinds of body-based experiences: egocentric information, i.e., relations of objects in the world to a static person viewing it, and inertial information generated when a person moves, e.g., a series of translations and rotations that can be used to form a route using path integration. World-based or allocentric knowledge for humans typically involves stable visual landmarks that can also form a route description or provide an overall framework. There is also available information from auditory, olfactory, or other senses, sometimes forming a gradient and other times providing more punctate location information.

Eventually, map-based representations may emerge for some individuals in some environments, although there is wide variation in whether they do (Peer et al. 2021). Environmental circumstances may favour certain strategies over others. For example, rural settings seem to support the development of use of distal cues and effective wayfinding, but city settings that occlude distal cues, especially if they include a regular street grid, are associated with less well-developed navigational ability (Coutrot et al. 2022; Hund et al. 2012).

How Are the Two Kinds of Memory and the Two Kinds of Navigation Related?

Buzsáki and Moser (2013) proposed a model of the relations among types of memory and types of navigation that they succinctly summarized in Figure 2.1. Although they meant to suggest contemporaneous relations and did not propose a developmental model, if one mental faculty supports another one, it would be natural to suppose that the more basic function comes first in development. However, contrary to the implication of Figure 2.1, semantic memory precedes episodic memory rather than the opposite. Path integration and allocentric representations develop in parallel, rather than path integration coming first. Semantic memory comes long before map-based navigation, as well as before episodic memory. Episodic memory and both forms of navigation develop largely in parallel. We review the four pairwise relations in the next four sections. (See Figure 2.2 for a developmental timeline.) In the concluding section, we briefly discuss how developmental facts can and should constrain our models of the mind and brain.

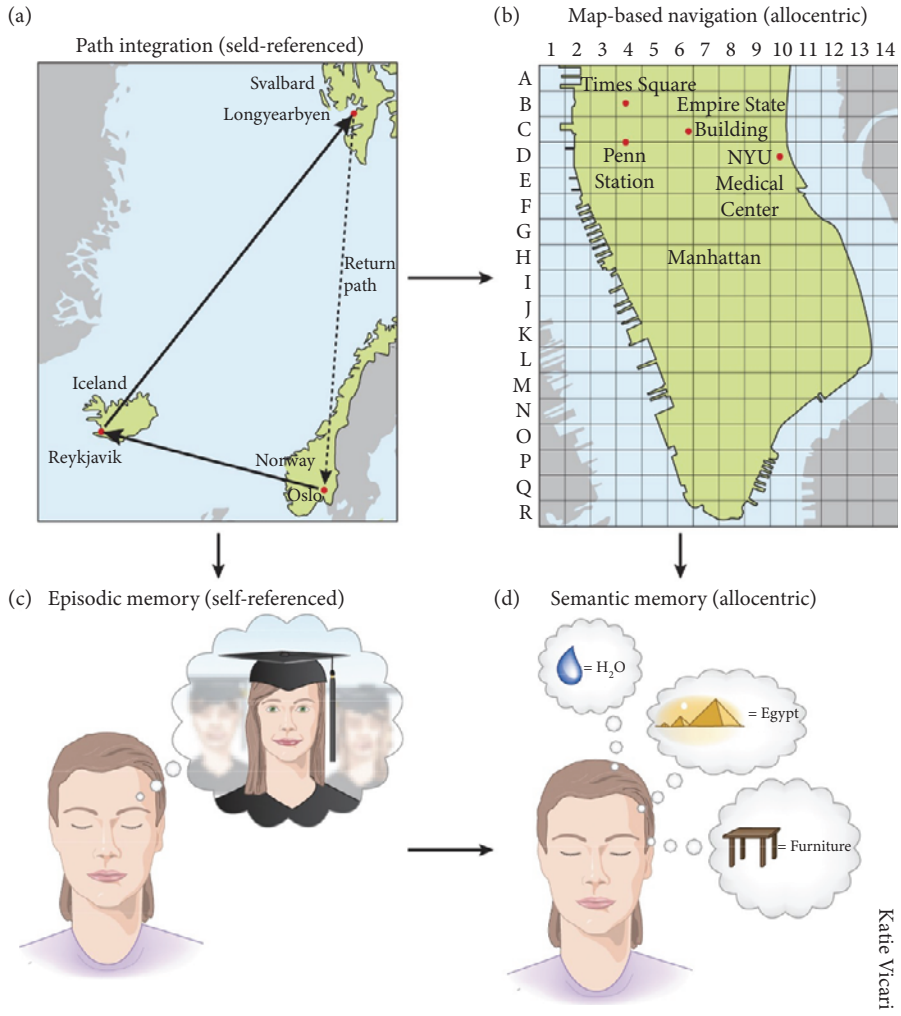


Figure 2.1 A proposed relationship between types of navigation and types of memory. Path integration features similar processes to episodic memory due to the cognitive effort to tie together parts of a whole. In path integration, two points may be known, and the navigator infers the third connection based on self-reference relations. Episodic memories are self-referenced and ease in recalling different features of the memory may prompt further details recalled. Map-based navigation includes relations among visible landmarks and a reconstructed map based on navigation experience. Semantic memory is the explicit and fact-based knowledge. Reproduced with permission from Buzsáki and Moser (2013), Springer Nature license #5704250345999.

Semantic Memory Before Episodic Memory

Many models of human memory suggest that semantic memory is based on abstractions from a series of related episodic memories (e.g., Kumaran & McClelland 2012). After repeated encounters with objects that enable sitting, sleeping, and eating in indoor spaces, people might form the concept of 'furniture'. After observing several

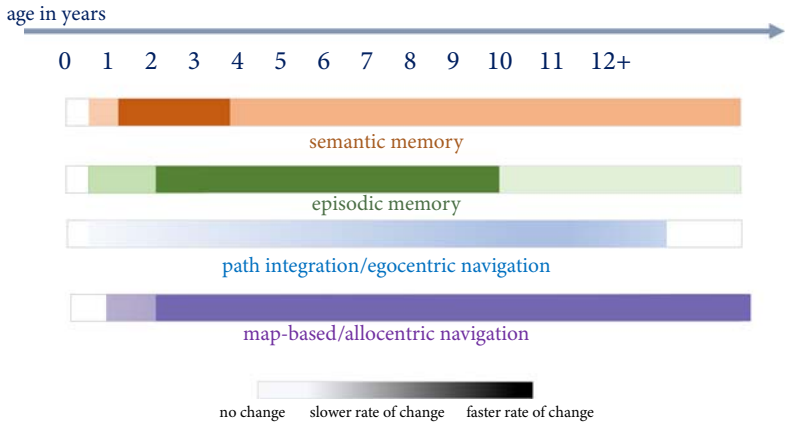


Figure 2.2 Developmental timeline of semantic memory, episodic memory, egocentric navigation, and allocentric navigation. The colour gradient indicates rates of change in learning. Children rapidly acquire semantic concepts in early childhood then learning is sustained through adolescence. Episodic memory follows a later emerging period of rapid learning. Both egocentric and allocentric navigation are observed early in infancy and follow longer lasting rates of change. However, egocentric navigation reaches adult-like levels earlier than allocentric navigation. CC-By Attribution 4.0 International from Nguyen and Newcombe (2024), osf.io/sntd2.

instances of a person painting or drawing, observers might conclude that the person is 'artistic'. Or after observing repeated interactions of people in fast-food restaurants, people might form a script for what happens in those contexts. Indeed, mature adults show linkages of this kind between episodic and semantic memory (e.g., Reagh & Ranganath 2023).

However, infants and toddlers do not learn about the world this way. They are rapidly acquiring concepts, categories, and vocabulary during the first two years of life, at a time when their episodic memories are (at best) fragile. Infants as early at 8 months can recognize statistical regularities of syllabic sounds in speech, with as little as 2 minutes of exposure (Saffran et al. 1996). By the end of the first year of life, babies comprehend many words and often produce a small number (Fenson et al. 1994). Over the next year, they rapidly learn to use hundreds of words (Frank et al. 2017). This learning through extracting regularities in speech gives rise to the rapid emergence of words, concepts, schemas, and scripts. In contrast, episodic memory during the first year is limited, with convincing evidence of contextually modulated memory not emerging until the second half of the second year (for review, see Newcombe et al. 2024). Most strikingly, older children and adults do not recall autobiographical memories from the first two years of life (Hayne & Jack 2011; Rubin 2000).

How do babies acquire concepts when their episodic memory is so limited? One prominent possibility involves statistical learning, in which infants can learn auditory and visual information by extracting statistical regularities of stimulus presentation. Prediction errors when viewing statistically predictable sequences

likely enable building event schemas (Wahlheim et al. 2022), and, even early in infancy, we can see the emergence of simple visual anticipation (Canfield & Haith 1991). The fact that young children often engage in repeated series of actions (e.g., bedtime routines) and enjoy repetition (e.g., repeated singing of the alphabet song) suggests the importance of this kind of learning. Associative and statistical learning at 6 and 10 months also predict greater vocabulary size at 18 months (Gerbrand et al. 2022).

Statistical learning may involve the monosynaptic pathway (MSP) relaying information from the entorhinal cortex (EC) to the hippocampal (HC) subfield cornu ammonis (CA) 1 and back to EC (Figure 2.3). In model simulation with a statistical learning paradigm, the slower inhibition rate and slower learning rate of the MSP better facilitated learning of regularities of overlapping representations (Schapiro et al. 2017). In contrast, the trisynaptic pathway (TSP) relays information from the EC -> dentate gyrus -> CA3 -> CA1 (Figure 2.3) and is vital for what-where-when associations, representing distinct episodes (Schapiro et al. 2017; Schlichting et al. 2017). Further, the MST can function independently of the TSP, possibly supporting infants' ability to acquire statistical information rapidly during infancy. CA1 develops before CA3, with a projection directly from entorhinal cortex (Lavenex & Lavenex 2013). Ellis et al. (2021) report data regarding the neural bases of early statistical learning. Infants from 3 to 23 months underwent fMRI scanning while viewing a stream of images with statistical structure. Hippocampal activation was mainly in the anterior HC, which may reflect the earlier-developing monosynaptic pathway. The babies also exhibited activation of medial prefrontal cortex, an area crucial for memory integration and generalization of information in adults (Raichle et al. 2001).

Statistical learning via the monosynaptic pathway cannot be the only possible route for the formation of semantic memory in the first years of life, however. First, consider the phenomenon of developmental amnesia (Elward & Vargha-Khadem 2018). Children who suffer hippocampal damage, often from anoxia at birth, show preserved ability to learn semantic information (Vargha-Khadem et al. 1997; Vargha-Khadem & Cacucci 2021). An especially dramatic example comes from a case report of a man who suffered exceptionally extensive HC damage, including damage to CA1 (Jonin et al. 2018). Despite this widespread lesion, he learned a great deal about the world. The famous patient HM could learn new semantic knowledge after bilateral MTL damage, albeit with some impairments (O'Kane et al. 2004). Second, we need to consider what kind of statistical learning is present in infancy (Forest, Schlichting et al. 2023). The indirect measures mainly used with infants, such as reaction time, looking time, and saccade latencies, show little developmental change (Choi et al. 2020), but direct testing measures, in which people must judge what they have seen or heard before, do change across childhood (Raviv & Arnon 2018). Is the knowledge reflected in indirect measures sufficient to support vocabulary growth and the building of scripts and routines? This idea needs evaluation.

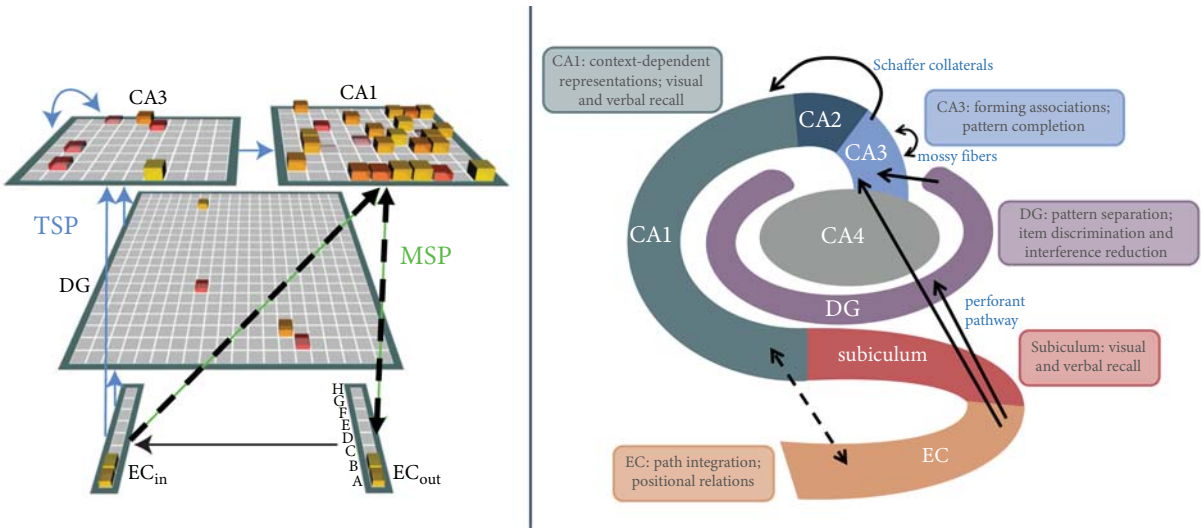


Figure 2.3 Hippocampal subfields and pathways. Left: The model architecture of the monosynaptic pathway (MSP) and the trisynaptic pathway (TSP) from Schapiro et al. (2017). The network model is trained to take entorhinal cortex inputs (EC_{in}) through both pathways to learn episodic content. The learned episode was successfully reproduced as EC output (EC_{out}). Reproduced with permission from Schapiro et al. (2017), CCC license #1549648-1. Right: Diagram of the HC subfields and associated proposed functions. Dotted arrow indicates the MSP, solid arrows indicate the TSP. CC-By Attribution 4.0 International from Nguyen and Newcombe (2024), osf.io/sntd2.

What other routes to semantic learning might supplement or even substitute for the monosynaptic pathway in building vocabulary and concepts? Research with developmental amnesia patients shows that repeatedly encountering material leads to them to develop a feeling of familiarity, as tapped by recognition testing. Familiarity might support the acquisition of semantic information (Elward et al. 2024) although as with indirect assessment of statistical learning, this idea requires evaluation. It is also possible, as Forest and her co-authors suggest, that infants may form very short-lived episodic memories that might support the development of gist memory, but the evidence for such memories is scanty (Forest, Schlichting et al. 2023).

The precedence of semantic over episodic memory is not confined to the first two years of life. Even preschoolers, who do form episodic memories, show less dependence on episodic memory for forming semantic memory than older children and in adults. Instead, young children depend on their existing semantic networks to support new generalizations. Ngo and colleagues (2021) measured generalization success in relation to context binding, item conceptual specificity, and item perceptual specificity. In adults, but not in young children, the ability to recall specific contexts promoted generalization, in line with the idea that episodic instances form the basis for a new semantic fact. In contrast, for young children, item conceptual specificity and the semantic linkages among instances were significantly related to successful generalization but episodic specificity was not. For slightly older children, none of the variables tested related clearly to generalization, raising the mystery of how transitions to mature episodic memory during elementary school relate to semantic memory. Consolidation through sleep may be an added factor to this relationship. After one night of sleep, older children had greater retention of generalized inferences than specific details of memories compared to younger children (Buchberger et al. 2024).

The generalizations studied by Ngo et al. (2021) involved attributes of characters; generalization was over contexts in which the characters were seen. However, a more powerful kind of generalization involves classes or categories. This situation also needs investigation. We do know that some kinds of generalization appear only in middle childhood, when we see combination across multiple memories to generate new inferences, as when self-deriving new semantic knowledge (Bauer et al. 2012, 2021), inferring novel associations between items indirectly related through a common link (Schlichting et al. 2017, 2022; Shing et al. 2019), or learning general rules in statistical learning paradigms (Forest, Abolghasem et al. 2023).

Overall, there are likely to be a variety of routes to concept formation, the acquisition of scripts, and generalization of various kinds. Some routes may be more utilized at specific developmental periods or in atypical development. In any case, episodic memory does not seem necessary for semantic knowledge acquisition and rapid generalization, although it is clearly sufficient for healthy young adults.

Intertwined Development of Path Integration and Map-Based Navigation

We turn next to the idea that path integration precedes map-based navigation. In rats, there is an established timeline for the emergence of the neurons that support navigation, beginning with head direction cells at postnatal day 12 (P12), followed by place cells, boundary cells and finally grid cells (Tan et al. 2017). Allocentric spatial abilities appear during this rapid period of development, maturing from P16 to P25 depending on the paradigm and the behaviour (Brown & Whishaw 2000; Shan et al. 2023). Thus, path integration (at least as picked up by the activity of head direction cells) does appear to precede allocentric navigation. However, drawing an inference about ordering of behavioural emergence in humans from the existing data on rats may be unwarranted, given that very little behavioural work has been done with developing rodents.

Nevertheless, it is striking that Piaget's observations of his own children suggested to him that egocentric spatial coding comes first in human development. Subsequent research has suggested that egocentric responding dominates during a period when babies have few motor capabilities and hence limited inertial information, and when their access to the visual world may also be limited by a visual system that lacks acuity for distal objects. Still, although at about 6 to 11 months infants often react to external stimuli in an egocentric manner, turning their heads to a previously learned cue (Acredolo 1978), in a landmark-rich room or familiar environment (such as their home), infants of this age can exhibit allocentric search (Acredolo 1979; Acredolo & Evans 1980).

Furthermore, egocentric coding is not the same as path integration, which develops in conjunction with the unfolding of motor milestones. Once infants become experienced crawlers around 8 or 9 months, they begin to code inertial information as they explore their environments (Adolph 2008; Adolph & Tamis-Lemonda 2014). They encounter obstacles and discern affordances while navigating around 12 months (Kretch & Adolph 2013). Crawling offers a limited field of view, however, so the onset of walking erect is another important milestone in inertial navigation. Novice walkers do not carry over the benefits of crawling, leading to poor judgment of affordances and consequent risk of falling and tripping (Adolph 2008; Kretch & Adolph 2013). Once children become strong walkers, they gain a wider field of view and can traverse greater distances, feeding into their developing navigation systems, both path integration and the allocentric system.

Path integration is often tested by having children traverse a path and then point or walk back to the starting position. Children as young as two years can path integrate without visual information (Landau et al. 1981; Rider & Rieser 1988). Four-year-olds can point back to the start position with comparable

accuracy as adults, although their orientation estimates are still lacking at this age (Rieser & Rider 1991). When blindfolded, 90% of a sample of 5- to 9-year-old children were able to path integrate back to a home position, but with less precision than adults (Bostelmann et al. 2020). Only 64% of these children could build an accurate cognitive map based on path information they learned without visual cues (Bostelmann et al. 2020). In a triangle completion task, blindfolded children from 5 to 7 years were still improving, not yet at adult levels, in heading direction and distance estimations to create a third, hypotenuse, vector connecting the two points (Smith et al. 2013). In short, it takes years after children can walk independently for them to hone their path integration system.

Allocentric navigation involves metric coding of distances, landmark use, and survey knowledge. By 16 to 24 months, children can use distance information to code object location independent of landmark or body position, although they do not seem to use categories to organize their environment into subsections until later (Huttenlocher et al. 1994). Children's metric coding accuracy continues to increase with age (Lambert et al. 2015; Simmering et al. 2008). From 6 to 12 years, children rely less on landmarks directly associated with turns (Jansen-Osmann & Fuchs 2006; Jansen-Osmann & Wiedenbauer 2004) and instead show flexible use of cues within the environment and distal cues (Buckley et al. 2015; Bullens, Iglói et al. 2010; Bullens, Nardini et al. 2010; Laurance et al. 2003). Using a proximal landmark is adult-like around 10 to 12 years, but using a boundary cue is not until 15 to 17 years (Glöckner et al. 2021).

The combination of sensory cues is vital to effective navigation. Around 9 years, children can optimally combine audio-visual cues (Negen et al. 2016), proprioceptive cues (Nardini et al. 2013), and metric information such as radius and angle (Sandberg et al. 1996). Children in middle childhood can use a combination of navigational frameworks and strategies to reconcile their goal-directed behaviour (Negen et al. 2016, 2019). There are also developmental changes in the ability to weigh and integrate the two systems. Cross-modal integration of spatial and sensory information seems to follow a long period of development behaviourally (Gori et al. 2008; Nardini et al. 2008, 2010) and neurally (Dekker et al. 2015), and the development of multisensory spatial processing also extends into early adolescence (Bruns & Röder 2023). By early adolescence we see adult levels of cognitive mapping (Bécu et al. 2020; Broadbent et al. 2014; Brucato et al. 2022; Bullens, Iglói et al. 2010; Glöckner et al. 2021; Nazareth et al. 2018; Nys et al. 2015).

Overall, the claim that egocentric spatial coding precedes allocentric coding has some limited validity in the opening sequence of both rat and human development. However, egocentric coding is not the same as inertial navigation or path integration. Path integration develops in tandem with allocentric coding, with a culminating challenge for map-based navigation being to appropriately integrate the two sources of information.

Semantic Memory Before Map-Based Navigation

Recently, some investigators have proposed that there are cognitive maps of world knowledge (e.g., Behrens et al. 2018). These discussions assume the existence of a Euclidean cognitive map, which is an idea that is hotly discussed among navigation researchers (see Peer et al. 2021), and extend it to a wide variety of kinds of flexible behaviour and the organization of systematic knowledge. They propose that map-like knowledge emerges from principles of reinforcement learning, and hence is not unique to the spatial domain.

However, there are reasons from thinking about developmental sequences to question this idea. First, as we have already discussed, semantic memory and organized knowledge emerge in the first few years of life, whereas cognitive maps (of the spatial variety) emerge much later, in early adolescence. Second, individuals differ substantially in their ability to form such maps, but there are no data suggesting that people who are navigationally challenged have difficulties with conceptual knowledge, and individual differences seem unlikely to be correlated. Third, individuals with developmental or adult-onset amnesia often show great difficulties in navigating and yet have good semantic knowledge.

One way to think about the hypothesis of cognitive maps of non-spatial knowledge is that conceptual knowledge can be spatialized by mature adults, at least after substantial effort. Most of the paradigms used in this line of research use very long training procedures, which not all participants succeed in completing. They also use knowledge domains that lend themselves to spatialization, e.g., arraying people or animals or objects along continuous dimensions. Thus, as with spatial cognitive maps, these conceptual maps likely depend on favourable circumstances and considerable effort.

Intertwined Development of Episodic Memory and Navigation

It is generally agreed that human navigation draws from memories of previous experiences and that episodic memory requires contextualization (Burgess et al. 2002; Ekstrom & Ranganath 2018). Several investigators suggest a direct link between episodic memory and navigation. For instance, working with epilepsy patients for whom intracranial electrodes can record neuronal firing, Miller et al. (2013) observed evidence of place cell activity both when participants were playing a taxi delivery game and during their subsequent free recall of locations visited. Of course, in this situation, the episodic task of free recall may be especially tightly linked to navigation, because most people follow a route as they engage in recall.

However, other investigators suggest only an indirect link between episodic memory and navigation, perhaps based on the common use of visuospatial imagery or scene construction in both tasks (Clark et al. 2019; Fan et al. 2023; Hassabis & Maguire 2007). Self-reported episodic memory and navigation form separate

cognitive clusters (Fan et al. 2021), perhaps suggesting very little overlap (Fan et al. 2023). Ekstrom and Hill (2023) point out that navigation requires sensory and visual cues and the combined weighting of cues to determine navigational significance and uses metrics that are not relevant to episodic memory such as distance estimations and visual/idiothetic integration (see Figure 2.4). Episodic memory is more internally driven than navigation, often self-referenced and not always tethered to spatial information (e.g., non-spatial associative memory).

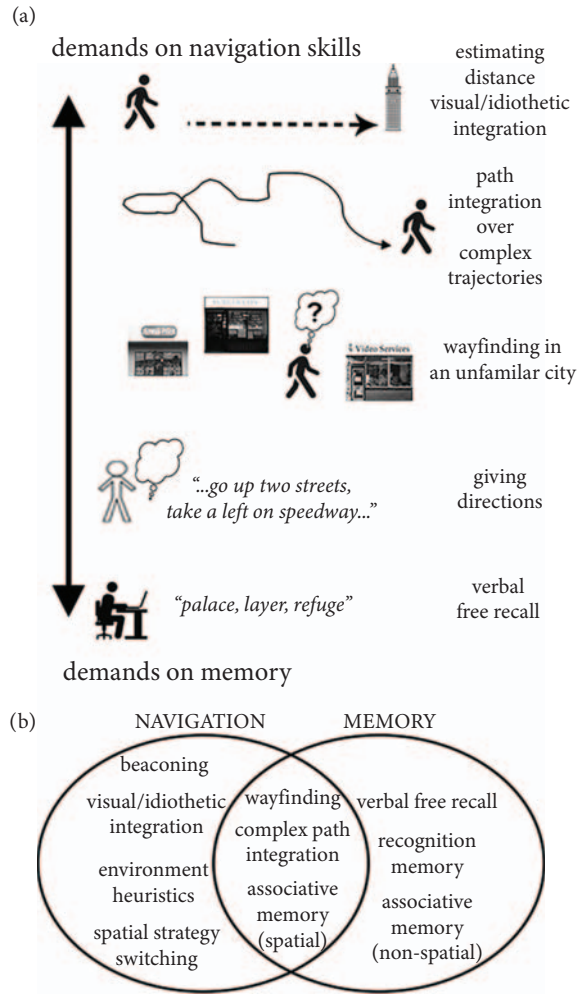


Figure 2.4 (A) Visualization of a continuous axis with the demands on navigation and memory on either end. The continuous model shows possible use to these systems in conjunction in some cases (e.g., wayfinding) and cases where memory and navigation are used uniquely (e.g., verbal free recall and estimating distance, respectively). (B) Venn diagram showing unique and overlapping functions of navigation and memory. Reproduced with permission from Ekstrom and Hill (2023), Elsevier license #5916050861458.

One argument for linkage between episodic memory and navigation, however, is that, in early childhood, there are striking similarities between the age at the very first appearance of episodic memory (Bauer 2007; Bauer et al. 2000; Newcombe et al. 2014), and when we first see place learning (Balcomb et al. 2011; Newcombe et al. 1998), both at about 21 months. What about later development? We have already seen that path integration and map-based navigation do not reach mature levels until early adolescence. We see the same extended timeline for episodic memory, although there is more substantial change from 2 to 8 years than from 8 years to adolescence (Figure 2.2).

A thorough review of the development of episodic memory is available in Newcombe et al. (2024) but here are some illustrative studies. Recalling an episode that includes who, what, and where (e.g., Sam went to the park with a picnic basket) involves relational binding, which improves significantly from ages 4 to 6 years (Lloyd et al. 2009; Ngo et al. 2018; Pathman et al. 2013; Riggins 2014; Sluzenski et al. 2004). Children also show holistic recollection by 4 years, in which retrieval of one pair of elements of the story (i.e., Sam and basket) is associated with retrieval of another pair (i.e., Sam and park). However, the degree of dependency increases into adulthood (Ngo, Horner et al. 2019). Precise memory for object details, often called pattern separation also improves between 4 and 6 years (Ngo, Lin et al. 2019); further, even 6-year-olds lag in establishing pattern separation for contexts (Benear et al. 2021; Ngo, Horner et al. 2019).

Just as cognitive mapping continues to improve into early adolescence, episodic memory continues to be refined (for an overview, see Ghetti & Fandakova 2020). Interestingly, in terms of the relation between episodic memory and navigation, and the idea that scene construction is a link between the two, Fandakova et al. (2019) found that higher neural specificity in scene-specific regions is associated concurrently with better memory.

Overall, the timelines for the development of episodic memory and the development of map-based navigation are close enough to support the possibility of important relationships, especially given a striking temporal coincidence at the end of the second year. Future work will need to probe the question more closely by looking at situations in which participants learn spatial layouts and acquire episodic memories in the same situation at the same time, looking for correlations of individual differences and common neural bases.

Developmental Sequences and Cognitive Architecture

Considering the development of semantic and episodic memory in juxtaposition with the development of navigation suggests that semantic memory is early and primary. Thus, understanding its unfolding and neural supports in the first two years of life is key to understanding the origins of knowledge. Path integration is largely paced by motor maturation but as children grow physically and

accumulate experience with wayfinding, they engage in an extended period of calibration, within each relevant sense, across sensory modalities and with the information derived from external landmarks. The earliest indications of what-where-when memory coincide strikingly with the earliest indications of place learning, and the development of these two functions has roughly the same timeline. However, although there may be overlap between them, notably because both require scene analysis and retention, there is currently little evidence that they should be grouped together. Rather, there may simply be some componential overlap, along the lines visualized in Figure 2.2.

This review of the currently available descriptions of development helps to evaluate the generality of several theories of cognitive functioning in adulthood. We have seen reason to question two popular hypotheses in current psychology and neuroscience: (1) the idea that semantic memory usually arises from abstraction over individual related episodes, and (2) the idea that spatial and non-spatial cognitive maps are the same entity. We have suggested that episodic memory and navigation are related by partial overlap of component functions, although we have yet to determine which ones. In highlighting a componential analysis of the relation between episodic memory and navigation, we come back to the dilemma with which we opened. Perhaps a construct like scene analysis is more basic to our cognitive architecture than are constructs such as episodic memory or navigation. Episodic memory and navigation both have adaptive functions, but they may not be the basic 'elements' that we need to guide us to deep scientific understanding.

References

Acredolo, L. P. (1978). Development of spatial orientation in infancy. *Developmental Psychology*, 14(3). https://doi.org/10.1037/0012-1649.14.3.224

Acredolo, L. P. (1979). Laboratory versus home: The effect of environment on the 9-month-old infant's choice of spatial reference system. *Developmental Psychology*, 15(6). https://doi.org/10.1037/0012-1649.15.6.666

Acredolo, L. P., & Evans, D. (1980). Developmental changes in the effects of landmarks on infant spatial behavior. *Developmental Psychology*. https://doi.org/10.1037/0012-1649.16.4.312

Adolph, K. E. (2008). Learning to move. *Current Directions in Psychological Science*, 17(3). https://doi.org/10.1111/j.1467-8721.2008.00577.x

Adolph, K. E., & Tamis-Lemonda, C. S. (2014). The costs and benefits of development: The transition from crawling to walking. *Child Development Perspectives*, 8(4). https://doi.org/10.1111/cdep.12085

Balcomb, F., Newcombe, N. S., & Ferrara, K. (2011). Finding where and saying where: Developmental relationships between place learning and language in the first year. *Journal of Cognition and Development*. https://doi.org/10.1080/15248372.2010.544692

Bauer, P. J. (2007). Recall in infancy: A neurodevelopmental account. *Current Directions in Psychological Science*, 16(3). https://doi.org/10.1111/j.1467-8721.2007.00492.x

Bauer, P. J., Cronin-Golomb, L. M., Porter, B. M., Jaganjac, A., & Miller, H. E. (2021). Integration of memory content in adults and children: Developmental differences in task conditions and functional consequences. *Journal of Experimental Psychology: General*, 150(7). https://doi.org/10.1037/xge0000996

Bauer, P. J., King, J. E., Larkina, M., Varga, N. L., & White, E. A. (2012). Characters and clues: Factors affecting children's extension of knowledge through integration of separate episodes. *Journal of Experimental Child Psychology*, 111(4). https://doi.org/10.1016/j.jecp.2011.10.005

Bauer, P. J., Wenner, J. A., Dropik, P. L., & Wewerka, S. S. (2000). Parameters of remembering and forgetting in the transition from infancy to early childhood. *Monographs of the Society for Research in Child Development*, 65(4), i–213.

Bécu, M., Sheynikhovich, D., Ramanoël, S., Tatur, G., Ozier-Lafontaine, A., Sahel, J. A., & Arleo, A. (2020). Modulation of spatial cue processing across the lifespan: A geometric polarization of space restores allocentric navigation strategies in children and older adults. *BioRxiv*. https://doi.org/10.1101/2020.02.12.945808

Behrens, T. E. J., Muller, T. H., Whittington, J. C. R., Mark, S., Baram, A. B., Stachenfeld, K. L., & Kurth-Nelson, Z. (2018). What is a cognitive map? Organizing knowledge for flexible behavior. *Neuron*. https://doi.org/10.1016/j.neuron.2018.10.002

Benear, S. L., Ngo, C. T., Olson, I. R., & Newcombe, N. S. (2021). Understanding relational binding in early childhood: Interacting effects of overlap and delay. *Journal of Experimental Child Psychology*, 208. https://doi.org/10.1016/j.jecp.2021.105152

Bostelmann, M., Lavenex, P., & Banta Lavenex, P. (2020). Children five-to-nine years old can use path integration to build a cognitive map without vision. *Cognitive Psychology*, 121. https://doi.org/10.1016/j.cogpsych.2020.101307

Broadbent, H. J., Farran, E. K., & Tolmie, A. (2014). Egocentric and allocentric navigation strategies in Williams syndrome and typical development. *Developmental Science*. https://doi.org/10.1111/desc.12176

Brown, R. W., & Whishaw, I. Q. (2000). Similarities in the development of place and cue navigation by rats in a swimming pool. *Developmental Psychobiology*, 37(4). https://doi.org/10.1002/1098-2302(2000)37:4<238::AID-DEV4>3.0.CO;2-J

Brucato, M., Nazareth, A., & Newcombe, N. S. (2022). Longitudinal development of cognitive mapping from childhood to adolescence. *Journal of Experimental Child Psychology*, 219. https://doi.org/10.1016/j.jecp.2022.105412

Bruns, P., & Röder, B. (2023). Development and experience-dependence of multisensory spatial processing. *Trends in Cognitive Sciences*, 27(10), 961–73. https://doi.org/https://doi.org/10.1016/j.tics.2023.04.012

Buchberger, E. S., Joechner, A. K., Ngo, C. T., Lindenberger, U., & Werkle-Bergner, M. (2024). Age differences in generalization, memory specificity, and their overnight fate in childhood. *Child Development*, 95(4), e270-86.

Buckley, M. G., Haselgrove, M., & Smith, A. D. (2015). The developmental trajectory of intramaze and extramaze landmark biases in spatial navigation: An unexpected journey. *Developmental Psychology*, 51(6). https://doi.org/10.1037/a0039054

Bullens, J., Iglói, K., Berthoz, A., Postma, A., & Rondi-Reig, L. (2010). Developmental time course of the acquisition of sequential egocentric and allocentric navigation strategies. *Journal of Experimental Child Psychology*. https://doi.org/10.1016/j.jecp.2010.05.010

Bullens, J., Nardini, M., Doeller, C. F., Braddick, O., Postma, A., & Burgess, N. (2010). The role of landmarks and boundaries in the development of spatial memory. *Developmental Science*. https://doi.org/10.1111/j.1467-7687.2009.00870.x

Burgess, N., Maguire, E. A., & O'Keefe, J. (2002). The human hippocampus and spatial and episodic memory. *Neuron*, 35(4). https://doi.org/10.1016/S0896-6273(02)00830-9

Buzsáki, G., Moser, E. Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nature Neuroscience*, 16, 130–138 (2013). https://doi.org/10.1038/nn.3304

Canfield, R. L., & Haith, M. M. (1991). Young infants' visual expectations for symmetric and asymmetric stimulus sequences. *Developmental Psychology*, 27(2). https://doi.org/10.1037//0012-1649.27.2.198

Choi, D., Batterink, L. J., Black, A. K., Paller, K. A., & Werker, J. F. (2020). Preverbal infants discover statistical word patterns at similar rates as adults: Evidence from neural entrainment. *Psychological Science*, 31(9). https://doi.org/10.1177/0956797620933237

Clark, I. A., Hotchin, V., Monk, A., Pizzamiglio, G., Liefgreen, A., & Maguire, E. A. (2019). Identifying the cognitive processes underpinning hippocampal-dependent tasks. *Journal of Experimental Psychology: General*, 148(11), 1861. https://doi.org/10.1037/xge0000582

Coutrot, A., Manley, E., Goodroe, S., Gahnstrom, C., Filomena, G., Yesiltepe, D., Dalton, R. C., Wiener, J. M., Hölscher, C., Hornberger, M., & Spiers, H. J. (2022). Entropy of city street networks linked to future spatial navigation ability. *Nature*, 604(7904). https://doi.org/10.1038/s41586-022-04486-7

Dekker, T. M., Ban, H., Van Der Velde, B., Sereno, M. I., Welchman, A. E., & Nardini, M. (2015). Late development of cue integration is linked to sensory fusion in cortex. *Current Biology*, 25(21). https://doi.org/10.1016/j.cub.2015.09.043

Dubova, M., & Goldstone, R. L. (2023). Carving joints into nature: Reengineering scientific concepts in light of concept-laden evidence. *Trends in Cognitive Sciences*, 27(7), 656–70.

Ekstrom, A. D., & Hill, P. F. (2023). Spatial navigation and memory: A review of the similarities and differences relevant to brain models and age. *Neuron*, 111(7), 1038–49.

Ekstrom, A. D., & Isham, E. A. (2017). Human spatial navigation: Representations across dimensions and scales. *Current Opinion in Behavioral Sciences*, 17. https://doi.org/10.1016/j.cobeha.2017.06.005

Ekstrom, A. D., & Ranganath, C. (2018). Space, time, and episodic memory: The hippocampus is all over the cognitive map. *Hippocampus*. https://doi.org/10.1002/hipo.22750

Ellis, C. T., Skalaban, L. J., Yates, T. S., Bejjanki, V. R., Córdova, N. I., & Turk-Browne, N. B. (2021). Evidence of hippocampal learning in human infants. *Current Biology*, 31(15). https://doi.org/10.1016/j.cub.2021.04.072

Elward, R. L., & Vargha-Khadem, F. (2018). Semantic memory in developmental amnesia. *Neuroscience Letters*, 680. https://doi.org/10.1016/j.neulet.2018.04.040

Elward, R., Limond, J., Chareyron, L. J., Ethapemi, J., & Vargha-Khadem, F. (2024). Using recognition testing to support semantic learning in developmental amnesia. *Neuropsychological Rehabilitation*, 34(8), 1141-60.

Fan, C. L., Abdi, H., & Levine, B. (2021). On the relationship between trait autobiographical episodic memory and spatial navigation. *Memory and Cognition*, 49(2). https://doi.org/10.3758/s13421-020-01093-7

Fan, C. L., Sokolowski, H. M., Rosenbaum, R. S., & Levine, B. (2023). What about "space" is important for episodic memory?. *Wiley Interdisciplinary Reviews: Cognitive Science*, 14(3), e1645.

Fandakova, Y., Leckey, S., Driver, C. C., Bunge, S. A., & Ghetti, S. (2019). Neural specificity of scene representations is related to memory performance in childhood. *NeuroImage*, 199. https://doi.org/10.1016/j.neuroimage.2019.05.050

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., & Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, 59(5). https://doi.org/10.2307/1166093

Forest, T. A., Abolghasem, Z., Finn, A. S., & Schlichting, M. L. (2023). Memories of structured input become increasingly distorted across development. *Child Development*, 94(5), e279–e295.

Forest, T. A., Schlichting, M. L., Duncan, K. D., & Finn, A. S. (2023). Changes in statistical learning across development. *Nature Reviews Psychology*. https://doi.org/10.1038/s44159-023-00157-0

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, 44(3). https://doi.org/10.1017/S0305000916000209

Gerbrand, A., Gredebäck, G., Hedenius, M., Forsman, L., & Lindskog, M. (2022). Statistical learning in infancy predicts vocabulary size in toddlerhood. *Infancy*, 27(4). https://doi.org/10.1111/infa.12471

Ghetti, S., & Fandakova, Y. (2020). Neural development of memory and metamemory in childhood and adolescence: Toward an integrative model of the development of episodic recollection. *Annual Reviews of Developmental Psychology*, 2, 365–88.

Glöckner, F., Schuck, N. W., & Li, S. C. (2021). Differential prioritization of intramaze cue and boundary information during spatial navigation across the human lifespan. *Scientific Reports*, 11(1). https://doi.org/10.1038/s41598-021-94530-9

Gori, M., Del Viva, M., Sandini, G., & Burr, D. (2008). Young children do not integrate visual and haptic information. *Nature Precedings*. https://doi.org/10.1038/npre.2008.1521.1

Hassabis, D., & Maguire, E. A. (2007). Deconstructing episodic memory with construction. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2007.05.001

Hayne, H., & Jack, F. (2011). Childhood amnesia. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(2). https://doi.org/10.1002/wcs.107

Hund, A. M., Schmettow, M., & Noordzij, M. L. (2012). The impact of culture and recipient perspective on direction giving in the service of wayfinding. *Journal of Environmental Psychology*, 32(4). https://doi.org/10.1016/j.jenvp.2012.05.007

Huttenlocher, J., Newcombe, N., & Sandberg, E. H. (1994). The coding of spatial location in young children. *Cognitive Psychology*, 27(2). https://doi.org/10.1006/cogp.1994.1014

Jansen-Osmann, P., & Fuchs, P. (2006). Wayfinding behavior and spatial knowledge of adults and children in a virtual environments: The role of landmarks. *Experimental Psychology*. https://doi.org/10.1027/1618-3169.53.3.171

Jansen-Osmann, P., & Wiedenbauer, G. (2004). The representation of landmarks and routes in children and adults: A study in a virtual environment. *Journal of Environmental Psychology*. https://doi.org/10.1016/j.jenvp.2004.08.003

Jonin, P. Y., Besson, G., La Joie, R., Pariente, J., Belliard, S., Barillot, C., & Barbeau, E. J. (2018). Superior explicit memory despite severe developmental amnesia: In-depth case study and neural correlates. *Hippocampus*, 28(12). https://doi.org/10.1002/hipo.23010

Kretch, K. S., & Adolph, K. E. (2013). Cliff or step? Posture-specific learning at the edge of a drop-off. *Child Development*, 84(1). https://doi.org/10.1111/j.1467-8624.2012.01842.x

Kumaran, D., & McClelland, J. L. (2012). Generalization through the recurrent interaction of episodic memories: A model of the hippocampal system. *Psychological Review*, 119(3). https://doi.org/10.1037/a0028681

Lambert, F. R., Lavenex, P., & Lavenex, P.B. (2015). Improvement of allocentric spatial memory resolution in children from 2 to 4 years of age. *International Journal of Behavioral Development*, 39(4). https://doi.org/10.1177/0165025415584808

Landau, B., Gleitman, H., & Spelke, E. (1981). Spatial knowledge and geometric representation in a child blind from birth. *Science*, 213(4513). https://doi.org/10.1126/science.7268438

Laurance, H. E., Learmonth, A. E., Nadel, L., & Jake Jacobs, W. (2003). Maturation of spatial navigation strategies: Convergent findings from computerized spatial environments and self-report. *Journal of Cognition and Development*, 4(2). https://doi.org/10.1207/S15327647JCD0402_04

Lavenex, P., & Banta Lavenex, P. (2013). Building hippocampal circuits to learn and remember: Insights into the development of human memory. *Behavioural Brain Research*, 254, 8–21. https://doi.org/10.1016/j.bbr.2013.02.007

Lloyd, M. E., Doydum, A. O., & Newcombe, N. S. (2009). Memory binding in early childhood: Evidence for a retrieval deficit. *Child Development*, 80(5), 1321–1328. https://doi.org/10.1111/j.1467-8624.2009.01353.x

Miller, J. F., Neufang, M., Solway, A., Brandt, A., Trippel, M., Mader, I., Hefft, S., Merkow, M., Polyn, S. M., Jacobs, J., Kahana, M. J., & Schulze-Bonhage, A. (2013). Neural activity in human hippocampal formation reveals the spatial context of retrieved memories. *Science*, 342(6162). https://doi.org/10.1126/science.1244056

Nardini, M., Bedford, R., & Mareschal, D. (2010). Fusion of visual cues is not mandatory in children. *Proceedings of the National Academy of Sciences of the United States of America*, 107(39). https://doi.org/10.1073/pnas.1001699107

Nardini, M., Begus, K., & Mareschal, D. (2013). Multisensory uncertainty reduction for hand localization in children and adults. *Journal of Experimental Psychology: Human Perception and Performance*, 39(3). https://doi.org/10.1037/a0030719

Nardini, M., Jones, P., Bedford, R., & Braddick, O. (2008). Development of cue integration in human navigation. *Current Biology*, 18(9). https://doi.org/10.1016/j.cub.2008.04.021

Nazareth, A., Weisberg, S. M., Margulis, K., & Newcombe, N. S. (2018). Charting the development of cognitive mapping. *Journal of Experimental Child Psychology*. https://doi.org/10.1016/j.jecp.2018.01.009

Negen, J., Ali, L. B., Chere, B., Roome, H. E., Park, Y., & Nardini, M. (2019). Coding locations relative to one or many landmarks in childhood. *PLoS Computational Biology*, 15(10). https://doi.org/10.1371/journal.pcbi.1007380

Negen, J., Roome, H., & Nardini, M. (2016). Young children can combine audio-visual cues near-optimally after training. *Journal of Vision*, 16(12). https://doi.org/10.1167/16.12.576

Newcombe, N., Huttenlocher, J., Drummey, A. B., & Wiley, J. G. (1998). The development of spatial location coding: Place learning and dead reckoning in the second and third years. *Cognitive Development*. https://doi.org/10.1016/S0885-2014(98)90038-7

Newcombe, N. S. (2018). Individual variation in human navigation. *Current Biology*, 28(17). https://doi.org/10.1016/j.cub.2018.04.053

Newcombe, N. S., Balcomb, F., Ferrara, K., Hansen, M., & Koski, J. (2014). Two rooms, two representations? Episodic-like memory in toddlers and preschoolers. *Developmental Science*, 17(5). https://doi.org/10.1111/desc.12162

Newcombe, N. S., Benear, S. L., Ngo, C. T., & Olson, I. R. (2024). Memory in infancy and childhood. In M. Kahana & A. Wagner (eds.), *The Oxford Handbook of Human Memory, Two Volume Pack: Foundations and Applications*. Oxford University Press. 1547–1575.

Newcombe, N. S., & Huttenlocher, J. (2000). Development of spatial thought. In *Making Space: The Development of Spatial Representation and Reasoning*. MIT Press. 109–144.

Ngo, C. T., Benear, S. L., Popal, H., Olson, I. R., & Newcombe, N. S. (2021). Contingency of semantic generalization on episodic specificity varies across development. *Current Biology*, 31(12). https://doi.org/10.1016/j.cub.2021.03.088

Ngo, C. T., Horner, A. J., Newcombe, N. S., & Olson, I. R. (2019). Development of holistic episodic recollection. *Psychological Science*, 30(12). https://doi.org/10.1177/0956797619879441

Ngo, C. T., Lin, Y., Newcombe, N. S., & Olson, I. R. (2019). Building up and wearing down episodic memory: Mnemonic discrimination and relational binding. *Journal of Experimental Psychology: General*, 148(9). https://doi.org/10.1037/xge0000583

Ngo, C. T., Newcombe, N. S., & Olson, I. R. (2018). The ontogeny of relational memory and pattern separation. *Developmental Science*, 21(2). https://doi.org/10.1111/desc.12556

Nguyen, K., & Newcombe, N. (2024). Developmental Sequences Constrain Models of the Mind figures. https://doi.org/10.17605/OSF.IO/SW4FD

Nys, M., Gyselinck, V., Orriols, E., & Hickmann, M. (2015). Landmark and route knowledge in children's spatial representation of a virtual environment. *Frontiers in Psychology*, 6(January). https://doi.org/10.3389/fpsyg.2015.00522

O'Kane, G., Kensinger, E. A., & Corkin, S. (2004). Evidence for semantic learning in profound amnesia: An investigation with patient H.M. *Hippocampus*, 14(4), 417–25. https://doi.org/10.1002/hipo.20005

Pathman, T., Larkina, M., Burch, M. M., & Bauer, P. J. (2013). Young children's memory for the times of personal past events. *Journal of Cognition and Development*. https://doi.org/10.1080/15248372.2011.641185

Peer, M., Brunec, I. K., Newcombe, N. S., & Epstein, R. A. (2021). Structuring knowledge with cognitive maps and cognitive graphs. *Trends in Cognitive Sciences*, 25(1), 37–54. https://doi.org/10.1016/j.tics.2020.10.004

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences of the United States of America*, 98(2). https://doi.org/10.1073/pnas.98.2.676

Raviv, L., & Arnon, I. (2018). The developmental trajectory of children's auditory and visual statistical learning abilities: Modality-based differences in the effect of age. *Developmental Science*, 21(4). https://doi.org/10.1111/desc.12593

Reagh, Z. M., & Ranganath, C. (2023). Flexible reuse of cortico-hippocampal representations during encoding and recall of naturalistic events. *Nature Communications*, 14(1). https://doi.org/10.1038/s41467-023-36805-5

Renoult, L., Irish, M., Moscovitch, M., & Rugg, M. D. (2019). From knowing to remembering: The semantic–episodic distinction. *Trends in Cognitive Sciences*, 23(12). https://doi.org/10.1016/j.tics.2019.09.008

Rider, E. A., & Rieser, J. J. (1988). Pointing at objects in other rooms: young children's sensitivity to perspective after walking with and without vision. *Child Development*, 59(2). https://doi.org/10.1111/j.1467-8624.1988.tb01482.x

Rieser, J. J., & Rider, E. A. (1991). Young children's spatial orientation with respect to multiple targets when walking without vision. *Developmental Psychology*, 27(1). https://doi.org/10.1037/0012-1649.27.1.97

Riggins, T. (2014). Longitudinal investigation of source memory reveals different developmental trajectories for item memory and binding. *Developmental Psychology*, 50(2). https://doi.org/10.1037/a0033622

Rubin, D. C. (2000). The distribution of early childhood memories. *Memory*, 8(4). https://doi.org/10.1080/096582100406810

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294). https://doi.org/10.1126/science.274.5294.1926

Sandberg, E. H., Huttenlocher, J., & Newcombe, N. (1996). The development of hierarchical representation of two-dimensional space. *Child Development*, 67(3). https://doi.org/10.1111/j.1467-8624.1996.tb01761.x

Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: A neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1711). https://doi.org/10.1098/rstb.2016.0049

Schlichting, M. L., Guarino, K. F., Roome, H. E., & Preston, A. R. (2022). Developmental differences in memory reactivation relate to encoding and inference in the human brain. *Nature Human Behaviour*, 6(3). https://doi.org/10.1038/s41562-021-01206-5

Schlichting, M. L., Guarino, K. F., Schapiro, A. C., Turk-Browne, N. B., & Preston, A. R. (2017). Hippocampal structure predicts statistical learning and associative inference abilities during development. *Journal of Cognitive Neuroscience*. https://doi.org/10.1162/jocn_a_01028

Shan, X., Contreras, M. P., Sawangjit, A., Dimitrov, S., Born, J., & Inostroza, M. (2023). Rearing is critical for forming spatial representations in developing rats. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4366412

Shing, Y. L., Finke, C., Hoffmann, M., Pajkert, A., Heekeren, H. R., & Ploner, C. J. (2019). Integrating across memory episodes: Developmental trends. *PLoS ONE*, 14(4). https://doi.org/10.1371/journal.pone.0215848

Simmering, V. R., Schutte, A. R., & Spencer, J. P. (2008). Generalizing the dynamic field theory of spatial cognition across real and developmental time scales. *Brain Research*, 1202. https://doi.org/10.1016/j.brainres.2007.06.081

Sluzenski, J., Newcombe, N. S., & Satlow, E. (2004). Knowing where things are in the second year of life: Implications for hippocampal development. *Journal of Cognitive Neuroscience*, 16(8). https://doi.org/10.1162/0898929042304804

Smith, A. D., McKeith, L., & Howard, C. J. (2013). The development of path integration: Combining estimations of distance and heading. *Experimental Brain Research*, 231(4). https://doi.org/10.1007/s00221-013-3709-8

Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory*, 82(3). https://doi.org/10.1016/j.nlm.2004.06.005

Squire, L. R., Knowlton, B., & Musen, G. (1993). The structure and organization of memory. *Annual Review of Psychology*, 44(1). https://doi.org/10.1146/annurev.ps.44.020193.002321

Tan, H. M., Wills, T. J., & Cacucci, F. (2017). The development of spatial and memory circuits in the rat. *Wiley Interdisciplinary Reviews: Cognitive Science*. https://doi.org/10.1002/wcs.1424

Tansan, M., Nguyen, K. V., & Newcombe, N. S. (2022). Spatial navigation in childhood and aging. *Annual Review of Developmental Psychology*, 4, 253–72. https://doi.org/10.1146/annurev-devpsych-121020-031846

Tulving, E. (1983). Elements of Episodic Memory. Canadian Psychology, 26(3).

Vargha-Khadem, F., & Cacucci, F. (2021). A brief history of developmental amnesia. *Neuropsychologia*, 150(107689). https://doi.org/10.1016/j.neuropsychologia.2020.107689.

Vargha-Khadem, F., Gadian, D. G., Watkins, K. E., Connelly, A., Van Paesschen, W., & Mishkin, M. (1997). Differential effects of early hippocampal pathology on episodic and semantic memory. *Science*, 277(5324). https://doi.org/10.1126/science.277.5324.376

Wahlheim, C. N., Eisenberg, M. L., Stawarczyk, D., & Zacks, J. M. (2022). Understanding everyday events: predictive-looking errors drive memory updating. *Psychological Science*, 33(5). https://doi.org/10.1177/09567976211053596

3

Space, Time, and Memory

György Buzsáki and János Végh

Time by itself does not exist . . . It must not be claimed that anyone can sense time apart from the movement of things (Lucretius, Book 1).

Introduction

Nothing is more intuitive, yet more complex, than the concepts of space and time. We can live without sight, sound, smell, touch, or even the ability to move, but everything we do feels to occur in space and time. These concepts are built in our language and thinking. For most cultures, space and time are used to explain the contrast between us as individuals against the vastness and complexities of the universe. Time and space are often used interchangeably ('my lab is 5 minutes from here'; Boroditsky 2018) and 'time' is among the most frequently used words in all cultures. Time is intimately related to movement and geometry. Scientific thinking has radically transformed these dimensionless concepts with the introduction of measuring instruments. Space and time were replaced with their definable variants: *distance/displacement* and *duration/interval*, which were quantified by the units of human-made instruments, such as rulers and clocks, thereby giving them practical meanings. As a result, the abstract concepts of space and time, defined as independent a priori from each other and from everything else (Kant 1998), have become constructed axioms of human thinking. The general appeal of space and time derives from the Newtonian framework of physics, referring to absolute time and absolute space, meaning that 'events' happen in a large 'theatre' (space) at some 'time' and any observer anywhere notices the same event at the same time. In our language-guided minds, time and space form the very basis of our imagination.

It would be hard to imagine and organize human life without clocks in the contemporary world. Time appears to be critical for all brain operations since every neuronal operation evolves 'in time', from coordination of muscle contractions during reaching to thinking and perhaps even dreaming (Tsao et al. 2022). Yet, when human-invented instruments are taken away, space and time revert back to dimensionless and ungrounded concepts. This happens because space and time, being immaterial, cannot exert an impact on things, including the body and the brain.

György Buzsáki and János Végh, *Space, Time, and Memory*. In: *Space, Time, and Memory*. Edited by: Lynn Nadel and Sara Aronowitz, Oxford University Press. DOI: 10.1093/oso/9780192882547.003.0003

How are then brains of animals expected to 'represent' space and time and their relationship when they have no sensors for either space or time, and no clocks or rulers to inspect? Furthermore, neither clocks nor brains 'make' time, only imagined concepts of such a thing (Buzsáki & Llinás 2017). Finally, since space and time cannot be studied directly, they cannot be derived from first principles. Yet, as we will discuss below, contemporary neuroscience research is still performed and interpreted largely within the 'theatre' framework of classical physics, even though in contemporary physics space is defined as the things themselves and there is no time 'in which' events occur.

As an alternative, we suggest that organization of brain dynamics offers a constraint how the brain interprets the relationships among objects and events out in the world. The hippocampus, a postulated key player in spatial navigation and memory, is the largest neuronal 'graph' in the brain, which constantly generates neuronal sequences or trajectories (Pastalkova et al. 2008). We will discuss that time and space are emergent properties of higher-order sequence learning (Buzsáki and Tingley 2019; George et al. 2021). The hippocampus is 'blind' to sensory modalities, concepts of space and time and, instead, performs a singular algorithm that learns a sequential, relational structure of its ecological niche without a need for spatial or time inputs or resorting to Euclidean assumptions. Geometric concepts emerge from brain dynamics and therefore, space (and time) should not be the task ('things-to-be-explained') for neuroscience.

What Happened to Me, Where, and When?

For the topic of this volume, time and space are special because of their postulated roles in episodic memory. Cognitive neuroscience distinguishes two types of memories that we can verbally declared: memories for events, referred to as episodic (or autobiographic) memory, and memories for facts, referred to as semantic memory (Squire 1992; Tulving 1972). Episodic events have a duration and take place somewhere. In contrast to this time-directed and segment-defined type of memory, declarable semantic facts are abstracted punctate events; they define objects, living things, and facts in the surrounding world. Semantic memories are gradually formed from multiple overlapping episodes with common items (junctions) among the episodes (Buzsáki 2005; Buzsáki & Moser 2013; Nadel & Moscovitch 1997) and through the repetitions the 'what' becomes invariant to the temporal and spatial conditions of the individual episodes (Figure 3.1).

The separation and independence of the of *what*, *where*, and *when* and their recombination for recall appear to be an efficient economic solution, compared to an alternative in which every individual experience of our lifetime was stored separately in an extraordinarily long list (Eichenbaum 2004; Tulving 1972; Squire 1992). An episode is an unfolding storyline in Newtonian space-time coordinates.

For recall, we recombine the trajectory or sequence of events by multiplying the marginals, akin to the hypothesized process of generating colour by multiplying different portions of red, green, and blue wavelengths (Friston & Buzsáki 2016). By hypothesizing separate coordinates for the *what*, *where*, and *when*, it appeared that neuroscience has identified a road map for uncovering the neurophysiological mechanisms of episodic memory (Eichenbaum 2014). By designating hippocampal neurons as 'place cells' (O'Keefe and Nadel 1978), superficial entorhinal cortical neurons as 'grid cells' and calling the same neurons as 'time cells' (Eichenbaum 2014; Istkov et al. 2011) gave an impetus to this research programme. Yet, despite its appeal, this programme is based on an outdated framework and space-time conceptualization. To illustrate this caveat, it is useful to recapitulate briefly the evolution of thinking about space-time issues in physics and mathematics.

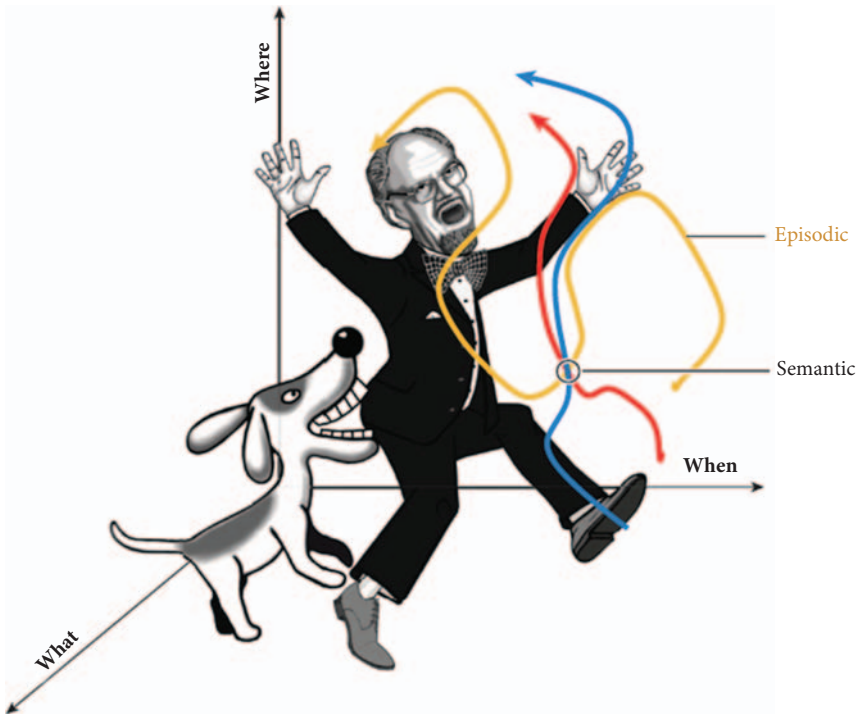


Figure 3.1 Space-time representation of episodic and semantic memories. Episodic memory evolves in space-time (what, where, when axes) with a corresponding neuronal trajectory (colour lines). The trajectory is a vector that unfolds in space-time (i.e., a segment with distance and duration). Only neuronal networks that generate such directional neuronal trajectories (as opposed to bidirectional correlations or associations) can support episodic memory. Three such trajectories (or episodes) are illustrated here. Intersection of multiple trajectories through the same space state can be considered as a node (black circle) or a snapshot in space-time (i.e., context-free semantic memory). Figure adapted from Buzsáki (2019).

Time of Physics and Time in the Brain

Modern science, including neuroscience, is quantitative: 'measure what is measurable, and make measurable what is not so' (https://www.goodreads.com/author/quotes/14,190.Galileo_Galilei). Measurements, in general, convert phenomena to numbers so that the observations can be written in the language of mathematics. In Newtonian physics, the objects on the scene of the theatre (i.e., space) interact instantly, assuming that the interaction speed is infinitely large. The Newtonian equations work reasonably well for describing the motion of the innermost planets of the solar system. However, if the interactions took time, that is if speed needs to be part of the model, the Newtonian model would no longer hold. It took two centuries to convincingly demonstrate that an observer can notice a distant event with a time delay instead of 'at the same time'.[1] From this new vantage point, modern physics did not invalidate the classic science. Instead, it drew the limits of the approximation it used and derived an approximation with extended range of validity (i.e., it offered a more comprehensive theory) by pointing out that the interaction speed must be considered explicitly.

Hermann Minkowski, Einstein's adviser in mathematics, wondered how the world would look like if the interaction speed between its objects were finite, although without specifying what the nature of the hypothetical limiting speed was (Pyenson 1977).[2] His model arose from experimental physics and through purely mathematical reasoning it lead to modified ideas about space and time.[3] The connection between physics and mathematics was the positive parameter *c*, which in the theory of special relativity was equated with the speed of light (Einstein 1905; Das 1993), emphasizing a finite speed of interaction. As in classical physics, where distance and duration are equated via speed and *c* is the key parameter in spacetime in the relativity theory. In contrast, neuroscience still considers space and time as orthogonal dimensions, even though speed has been shown to be an important gain factor in neuronal computation (Buzsáki 2019). Space and time can refer to processes in the brain but also the animal's actions. Speed in the brain has two components which relates to computing and transfer. Perhaps it is justifiable to entertain the idea that neuroscience can also exploit the Minkowski's model by examining the critical role of *c*, such as conduction speed of axons and locomotion speed (as will be discussed below), at every level of neuroscientific inquiry, including the investigation of memory.

[1] Experienced anomalies, for example deviations from the 'ideal' laws of motion of celestial bodies, did not immediately lead to revisiting the model. Based on timing the eclipses of the Jovian moon Io, Rømer suggested that the speed of light is finite (https://www.amnh.org/learn-teach/curriculum-collections/cosmic-horizons-book/ole-roemerspeed-of-light). But deviations rarely result in the rejection of a model. Only replacing the old with a more comprehensive model propels science forward.

[2] Minkowski's model was a three and a half dimensional coordinate system, since the four coordinates were linked through the interaction speed (as an external parameter for the model; Minkowski 1908).

[3] Eucledian geometry was no longer adequate to the task of describing physical reality (only an approximation), and had to be replaced by the geometry of a four-dimensional 'world.' 'Beyond the divisions of time and space which are imposed on our experience, there lies a higher reality, changeless, and independent of observer' (Galison 1979).

Computing, Processing Neuronal Information, and Time Delays

The idea that the brain computes can be traced back to the computer metaphor of von Neumann (1993). As result, 'computing', 'processing', and 'information' have become frequently used words for describing observations in neuroscience experiments, although without a consensus what these terms actually refer to (Krakauer et al. 2017).

Von Neumann suggested that a computing element must have an 'input section' and an 'output section'. In the case of chained (serial) operations, the data must be first transferred to the input section of one unit before performing an operation, and after the operation ('transformation' or 'abstraction' in neuroscience) the results are transferred from its output section to the input section of the next computing unit. All these changes evolve along a trajectory, where the rate of change can be conveniently compared to a reference clock time. In this formulation, the total operating time is separated into transfer time and processing (or computing) time (Mizuseki et al. 2011). In his classic computing paradigm, von Neumann omitted the transfer time but warned that with the changing technology ('using much faster vacuum tubes') the validity of that omission must be revisited. Further, he explicitly warned that applying his simplified paradigm 'to neural computing would be unsound' (*sic*; von Neumann 1993). However, he did not provide a solution for the general case of computing when the transfer time component not only cannot be omitted but even can dominate the total computing time, as is the case for the brain.

In the theory of special relativity all coordinates are distances and time is transformed to distance using the propagation speed of light, therefore the abstracted coordinate system is called 'space-time'. In contrast, in modern electronics and in the brain, the primary feature is time (duration) and distances are measured along the path in which the signal is transformed to time, therefore, it can be called time-space system. Transfer in any kind of computing needs finite time because of the finite interaction speed in the implementation (Végh 2020, 2021; Végh and Berki 2022). The implication of time delays for time-space can be illustrated by a general computing process in the 4-D coordinate system, an extension of the famous 'light propagation' Gedankenexperiment (http://visualrelativity.com/LIGHTCONE/minkowski.html), supplemented with the need for a finite time to switch the light on (Figure 3.2). In the model, the light is emitted in the origin (the first computing unit) and the observer (the second computing unit) is located at the position marked by a circle. After the light is turned on, the second computing unit must wait until after 'future light cone' reaches observer's position. The arrows at positions 1 and 2 show the timing relations for interaction speed differing by a factor of 2 and ½, respectively.

The constraint (or perhaps evolutionary advantages) of temporal delays in neuronal computation is a poorly investigated territory of neuroscience. Just two hundred years ago, the vitalist thinking claimed that the speed of the mind is

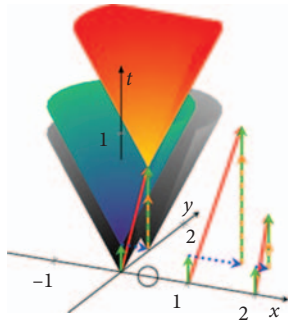


Figure 3.2 Critical importance of time delays in computation. The total computing time comprises components transfer and processing time. The processing occurs 'in place' (horizontal blue arrows) and the transfer between computing objects occurs 'in time' (vertical green arrows). The two operations mutually annihilate each other (mixed colour arrows), introducing inherent inefficiency into computing. The Minkowski-distance (red arrows) between the beginning and the end of computation differs for different signal propagation speeds. The computing time is a linear function of neither the transfer speed nor the processing speed. Reproduced from Végh (2021).

infinitely fast or, if it has a speed, it cannot be measured ('*ignoramus et ignorabimus*'; Muller 1840), but soon von Helmholtz have demonstrated that the propagation of electricity in the nerve is in fact quite slow (von Helmholz 1850). Subsequently it was established that the propagation speed of the action potential varies orders of magnitude in axons from tens of millimeters to meters/second, obeying a lognormal distribution (Buzsáki and Mizuseki 2014). An additional but small delay takes place in the chemical synapse (~0.3 ms conversion time from electrical to chemical signalling; Somjen 1972; Johnston and Wu 1994), leading to an analogue computation, measured as depolarization or hyperpolarization of the target membrane segment (Eccles 1957). A neuron integrates spike inputs from multiple upstream neurons in an analogue manner and reconverts them into a digital pulse code (action potentials or spikes) in the axon initial segment (Stuart et al. 1997). The transmission speed from the action potentials of upstream neurons to a downstream target neuron ('spike transmission delay') can vary substantially from a few milliseconds to hundreds of milliseconds, depending on the instantaneous conductance of the integrating neuron and the simultaneously acting feed-forward inhibition (Pouille and Scanziani 2001). Thus, while the axon conduction speed is constant, the efficacy of spike transmission varies across neurons and can vary as a function of the interaction of the neuron and the circuit in which the neuron is embedded. Therefore, the delay between Von Neumann's 'input section' and 'output section' can vary, in contrast to digital computation. Such time-varying computation may be 'most efficient in its use of resources is neither analogue computation nor digital computation but, rather, a mixture of the two forms' (Strong et al. (1998). This distinction was already foreseen by von Neumann: 'The language of the brain not the language of mathematics' (von Neumann 2012).

The time delays in brain computation also impacts how the brain handles information. The mathematical framework introduced by Claude Shannon (1948) for

information transfer is connected to physical reality through an arbitrary external parameter, the 'allowed signals of duration *T*' (aka, channel capacity). In Shannonian information theory interactions occurs at the speed of light and in practice the mathematical handling corresponds to instantaneous interactions. Applying information theory to brain computation with its millions of times lower interaction speed and variable delays is therefore hard to justify (Shannon 1956; Végh and Berki 2022). Any communication, including electrical or neuronal, needs a carrier, such as an electrical impulse or neural spike. Because this carrier has a temporal component, the implication is that the information itself has a propagation speed. The speed of information propagation in the Shannon model is that of the electromagnetic wave and the information density is defined by the signal frequency. In contrast, in neural communication a spike alone only represents the beginning of a new message (as a synchronizing signal, it 'wakes up' the receiver), without any hint what 'information' is being transmitted. Single spikes are not able to deliver information, in line with the information theoretical conclusion that 'if a source can produce only one particular message its entropy is zero' (Shannon 1948). Instead, neurons use temporal intervals to communicate, and logical operations (information processing) in an analogue manner. Next, we will discuss how these considerations apply to neuronal computation.

Time Is Neuronal Space in the Brain

'*Architecture makes music stand still. Here time turns into space*' (Robert Kelly: 'Phases of the Earth'; https://www.visibleweather.com/home/tspace.html). This poetic formulation of the relationship between space and time applies to the relationship between distance metric and duration metric in the brain. Simultaneity of two (or more) events may be deemed synchronous (i.e., occurring within a defined time interval of an observer) even if the two events occur at vastly different times. For example, action potentials arriving at the same time onto the dendrite of a reader neuron from a nearby and distant neuron exert a cooperative impact on discharge of the reader (target) neuron, even though the spikes in the two upstream neurons were generated tens of milliseconds apart (Figure 3.3A). Conversely, action potentials that are generated at the same (clock) time in a nearby and distant neuron, will arrive to the reader neuron tens of milliseconds apart (i.e., asynchronously; Figure 3.3B). This observer or 'reader'-defined synchrony is critical in brain operations. If the action potentials from many upstream neurons arrive within the membrane time constant of the target (reader) neuron (τ: 10–50 ms for a typical pyramidal neuron), their combined action is cooperative because each of them contributes to the discharge of the reader neuron. Action potentials arriving later can only contribute only to initiating another action potential. Thus, from the reader neuron's point of view, upstream partners that contribute to its spike discharge constitute a functional assembly (integrated by the membrane time constant), whereas spikes outside this time window can only be part of another assembly (Harris et al. 2003). This simple mechanism

can thus both integrate and segregate upstream neurons into discrete assemblies, irrespective of whether they are interconnected or not. The entire goal of synchronous cooperation is to have an impact, i.e., trigger an action potential in target neurons. Neurons which achieve this joint impact can be considered as a fundamental assembly. When the reader mechanism can integrate over longer time scales (e.g., NMDA channels or neuronal circuits), several fundamental assemblies can be concatenated.

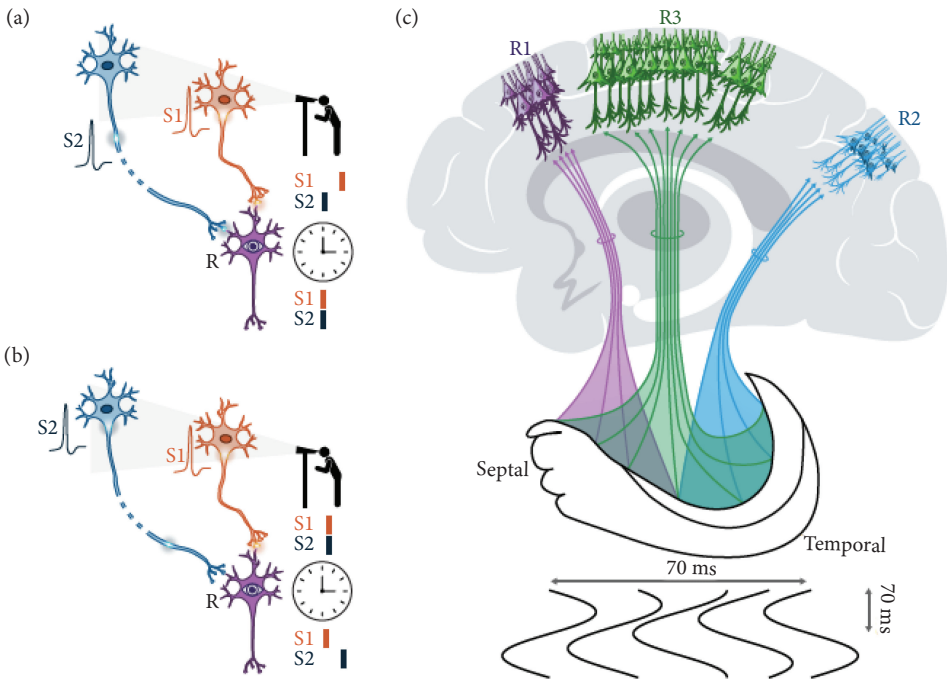


Figure 3.3 (A–C) Time is neuronal space. (A–B) Observer/reader-dependent definition of synchrony. (A) Spikes from upstream neurons (Sender 1 and Sender 2, S1, S2) arrive 'at the same time' to the dendrites of a reader neuron (R) but these spikes were initiated at different times by S1 and S2. (B) Spikes initiated at the same time in S1 and S2 arrive at different times to the Reader. The Reader-centric classification of spikes generalizes to populations as well. (C) Space-duration readout of neuronal messages. Spikes from the hippocampus that arrive within the time constant of the neocortical reader circuits R1 or R2 are integrated. Upstream (hippocampal) neurons which send these spike messages are considered separate assemblies by R1 and R2, respectively, due to timing of spikes and anatomical connectivity. The hypothetical reader circuit R3 has a 'preformed anticipatory schema' with a shared cipher (theta cycle frame) that helps to select features that are characteristic for anticipated events or activities and their change. Marking the message boundaries by cyclic inhibition partitions a continuous stream of events into discrete units and forms the hypothetical basis for extraction and synthesis of new information (a process called 'abstraction'). In the example, the message sent from the entire hippocampus occurs in half a theta cycle (~70 ms). Reprinted from Buzsáki and Voroslakos (2023).

This reader-centric view becomes of utmost importance when one wants to understand communication across brain regions, such as the hippocampus and neocortex (Figure 3.3C). Because of the slow communication across neurons via the relatively slow conducting axons and charge time of the neuronal membrane, computation and messaging are not instant but protracted over time. This slow propagation of the activity within and across networks is reflected at the mesoscopic level as 'traveling waves' (Ermentrout and Kleinfeld 2001). For example, population activity travels from the dorsal to the ventral pole of the hippocampus in half a theta cycle (~70 ms; Figure 3.3C). Therefore, the question that arises when the entire hippocampus takes part in a particular computation is: how does the neocortical reader integrate (i.e., decode) neuronal messages from the hippocampus? As is the case in human language communication, where even the last word of the sentence can change the meaning of the sentence, the reader structure should know both the beginning and end of each message—in our example, the spikes from the entire hippocampal volume within 70 ms timeframes. Separation of networks into senders and receivers serves only didactic purposes. In the complex networks of the brain, most structures are bidirectionally connected and can serve as both senders and receivers, which functions can change rapidly by shifting the phases of the sender and receiving partners.

In summary, communication across multiple neuronal circuits occurs in a time-evolved manner. The term communication refers to an agreement between a sender and receiver. This 'agreement' is called the cipher known to both parties. A key aspect of the cipher is a method by which the messages are broken down into smaller information chunks or segments, the *modus operandi* by which the sent messages are transcribed and translated by the receiver. Both the sent and received messages must have a start and end time point and in these intervals a large neuronal 'space' (or volume) is involved. Chunking of messages by agreed rules allows the generation of and reading (i.e., 'de-ciphering') virtually infinite combinations from a finite number of elements in human, sign, body, artificial, and computer languages, music, and mathematical logic, and, presumably, in the brain. This small set of rules, usually referred to as syntax, governs the combination and temporal progression of discrete elements (such as letters or musical notes) into ordered and hierarchical relations (words, phrases, and sentences, or chords and chord progressions) that allow the generation of messages, which when interpreted by the receiver *becomes* information. Without a reader-interpreter neuronal 'information' does not have a meaning. Since messages are sent to multiple targets with varying distances and often distinct internal dynamics, the same exact action potential trains can be interpreted differently by the distinct readers. The potential substrate for a hypothetical neuronal syntax in the brain is the constellation of brain rhythms (Buzsáki 2010). The oscillation cycles serve as the cipher, marking the beginning and end of messages from short 'neuronal letters' (Harris et al. 2003), which are concatenated to become 'neuronal words' by slower rhythms by cross-frequency phase-amplitude coupling (Buzsáki and Draguhn 2004). The coordinating role of oscillation cycles

across brain regions assures that messages from multiple sources arrive at the 'same time' (phase), despite variable delays via multiple paths (Sirota et al. 2008).

Distance to Time Compression via Speed

In the rodent hippocampus, neuronal activity is organized by a 6-10 Hz theta oscillation (Buzsáki 2002) and its pyramidal neurons are believed to 'respond' to particular constellation of the environment ('place cells'; O'Keefe and Nadel 1978). The place fields of place cells are typically several times larger than the length of the rat, and place fields of several neurons overlap with each other over multiple theta cycles during ambulation. In addition to this locomotion-related overlap at the seconds scale, several place cells fire together in a given theta cycle such that the spike timing sequence of neuronal assemblies predicts the sequence of passed and upcoming locations in the rat's path, with larger time (phase) lags between spiking of place cells within the theta cycle representing larger distances (Figure 3.4A–C; Skaggs et al. 1996; Dragoi and Buzsáki 2006; Diba and Buzsáki 2008; Bouffard et al. 2018). In other words, if we take a 'snapshot' over a single theta cycle, the spike sequences (or 'neuronal trajectories', Figure 3.1) correspond to the trajectory of place fields the animal has just passed and is going to visit ('spatial segment' or distance; Figure 3.4). One interpretation of this relationship is that distances in the world are transformed to durations in the brain ('space-time' of physics) on the assumption that the phase of spiking of hippocampal neurons are 'driven' by some external cues. A problem with this interpretation is that the theta rhythm is induced in the brain and its phase varies independently from the external cues so that upon repeated runs the phase of theta varies randomly relative to the same spatial locations.

An alternative interpretation is that the primary mechanisms is an internally maintained dynamics in the hippocampus, and this dynamics constrains how events in the world are matched to this pre-existing dynamics ('time-space'). In support of the latter interpretation, the time (theta phase) offsets in the hippocampus remain similar in different size environments, so that the same time segments (duration) between place cell spikes now correspond to larger distances in larger environments (Diba and Buzsáki 2008). From this perspective, the hippocampus can be viewed as a time-space zoom in which the neuronal resources are determined by an internal dynamics (the number of cell assemblies and the total number of spikes within theta cycles remain the same; Figure 3.4), therefore the resolution of the environment scales with its magnitude (in larger environments, the spatial resolution is poorer, with larger place fields and larger distances between place fields; Diba and Buzsáki 2008; Kjelstrup et al. 2008). What is the conversion mechanism in the brain between distance and duration?

In classical physics, distance and duration are equated via speed. One can expect that this general law also applies to organisms moving through space. As just discussed, each theta cycle contains a segment of travel (distance) coded as a sweep of

time (duration). How can this distance—duration relationship remain faithful when the animal runs at different velocities? If the rat traverses the place field of a neuron in 1 second during the first pass and then in half a second in the next pass, the place cell will be active for 8 and 4 theta cycles, respectively (assuming 8 Hz theta frequency). An important experimental observation is that the number of spikes within the place field remains the same even though the rat's velocity changes (Figure 3.4). For this reason, the number of spikes per theta wave and the overlap between spiking of successive place cells approximately doubles. Due to this speed-related firing rate gain per theta cycle, the magnitude of the cycle-to-cycle phase advance of spikes increases proportionally. Thus, speed gain compensates for the shorter time spent in the place field, leaving the relationship between theta phase and spatial position relatively invariant (Figure 3.4). Because the brain has constant access to velocity from the body and the vestibular system, time and distance travelled can be interchangeably calculated. Thus, the distance-to-duration compression may be only an appearance. Instead, the internally organized cell assembly sequences are available to match the animal's position to occasional landmarks but the occurrence of theta phase related spiking is not micromanaged by environmental cues but are controlled by internally generated dynamics.

The primacy of phase versus clock time (i.e., time related to a human-made device) is further supported by experiments in which the medial septum (a key structure that provides a 'pacemaker' theta signal to the hippocampus; Petsche et al. 1962) was cooled artificially to decrease the frequency of theta oscillation (Petersen and Buzsáki 2020). Cooling affected the distance-to-time, but not the distance-to-theta phase, conversion (Figure 3.4D, E). These findings revealed that cell assembly sequences in the hippocampus are not related to the time of reference clocks but to the phases of internally generated rhythms. In other words, brain 'time' and time measured by clocks, while they can be related, are independent.

Place Cells, Time Cells, Episode Cells, Tone Cells, and Many Other Names

A recent attempt to explain the role of time and relate it to space in cognitive phenomena is to distinct neuronal 'representations' of the future and the past (Tsao et al. 2022). According to this model, durations that start in the present and end in the future ('prospective timing') are supported 'clock-like mechanism that accurately tracks the continuing passage of time', either periodic, such as brain oscillations or cumulative. In contrast, durations that start in the past and end either in the past or the present ('retrospective timing'; 'episodic timing' or 'remembered duration') is 'computed only from memory of the interval and events occurring within it' (Tsao et al. 2022; Friedman 1993; Ornstein 1969). Thus, retrospective timing is 'the conversion of memory representations into a duration value' (Tsao et al. 2022). The main argument in favour of two distinct psychological times is that

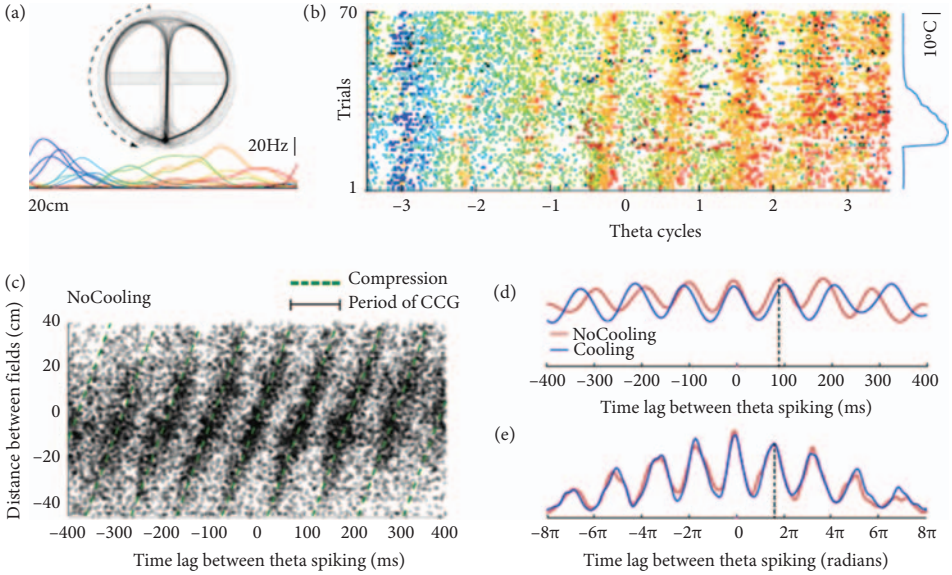


Figure 3.4 Distance-Theta Timescale Conversion. (A) Selected place fields in the left arm of the maze before medial septum cooling. Each coloured line is shows the smoothed firing rate of a place cell. (B) Within-theta cycle cell assemblies. Each row of dots is a trial of the spiking activity of place cells (same colour code as in (A) in successive theta cycles), organized into subsequent theta cycles, rather than distance or time. 0 is the theta cycle in the middle of the left arm. Note that cell assembly sequences within theta cycles (from dark blue to red) correspond to the order of place fields on the track. Note also that the phase preference of place cell spikes are preserved during medial septum cooling, (right trace), which manipulation decreased the oscillation frequency of the entire population. (C) Relationship between distances of place field peaks across neuron pairs (y axis) as the rat travels in the maze and their theta time-scale cross-correlogram lags (time lag on x axis) during control trials. Green dashed line; distance-time compression index (slope), c = 0.54 cm/ms. Note the reliable relationship between place field distances and time offset of spikes at theta (~100 ms) time scale. (D) Sum of the theta repeated spike cross-correlograms of place field pairs. The red line is the sum of all dots in (C) (NoCooling); the blue line is the sum of all dots of a similar graph during cooling. Vertical dashed line indicates that the oscillation frequency of the spike cross-correlation before cooling (~85 ms) is faster that the frequency of the population theta (~110 ms at 9 Hz). Note also that this relationship is increased during cooling. (E) Similar to (D), but instead of time lags, within-theta phase lags between spikes were calculated. Vertical dashed line indicates that the oscillation frequency of the spike cross-correlation is faster than the frequency of the population theta (2π). Note that the theta phase of spiking remained unaltered by slowing the theta frequency. After Petersen and Buzsáki (2020).

remembered time (retrospective) corresponds to 'significantly different duration estimates for the same physical time intervals' and the recall process is assisted by event boundaries that serve as time stamps or 'nows' (i.e., time points; Tsao et al. 2022). Unfortunately, this approach does not clarify what 'tracking of elapsed' time

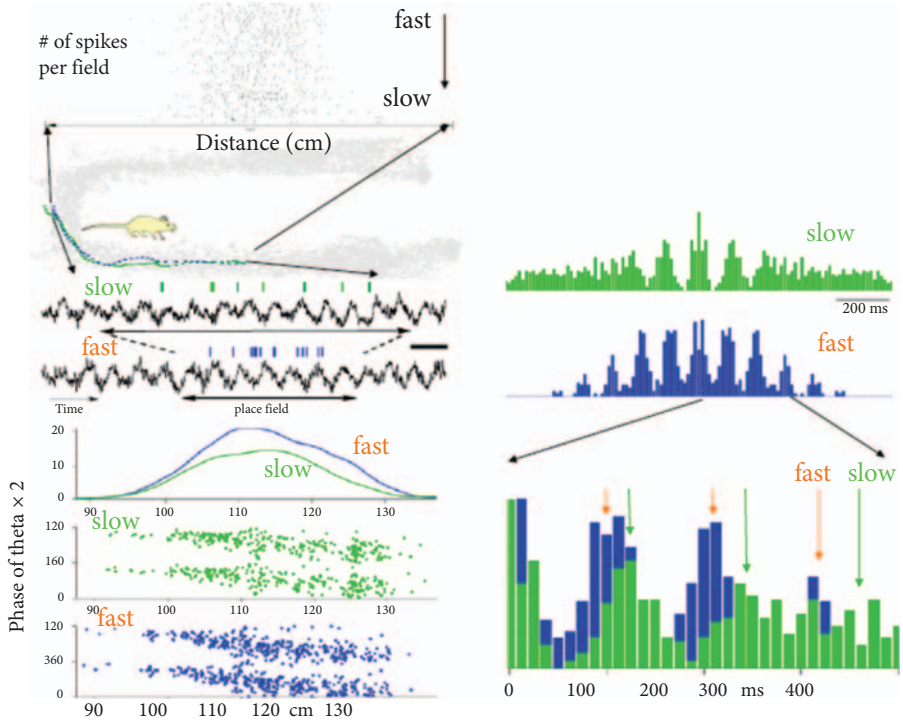


Figure 3.5 Speed-gain for distance-time compression in the hippocampus. (A) Top, The number of spikes (dots) within the neuron's place field is similar on slow and fast run trials. Trials are sorted by velocity from slowest to fastest. Middle, Movement trajectories of a rat through a place field on two trials with different speeds. Spikes (vertical ticks) and the corresponding theta rhythm of the same two trials (colour code). Horizontal arrows indicate the time it took for the rat to run through the place field. (B) Top, firing (place) field of the neuron during slow and fast runs (top half and bottom half speeds). Middle and bottom, Spike distributions within the place field as a function of the phase of the theta cycle during slow and fast runs. Note that the slope of the phase shift within the field remained the same, indicating that the position versus theta phase-spike relationship is preserved despite varying speed of the rat. (C) Oscillation frequency (shown here as spike autocorrelograms) of the same neuron during slow and fast trials. Note that the oscillation frequency of the place cell is faster during the faster trials. The duration x speed then is responsible for the preserved relationship between distance from the beginning of the place field and the rat's position. After Geisler et al. (2007).

by brain mechanisms actually mean. The tacit assumption is that through its sensors the brain can read genuine temporal clues and units that exist 'out in the world'. Unfortunately, there are no genuine *temporal* clues to read out there. Sunrise is a reliably recurring event that allows living things to anchor their biological processes but in itself is neither a 'time point' or duration. For retrospective timing, in contrast, the brain has to resort to reading its own dynamics without reference to outside events. Yet, it is hard to justify the assumption of distinct mechanism and a separation of

temporal phenomena. If the brain can generate its own time, why is there a need to postulate tracking of external time? Such attempts follow the dominant trends in current neuroscience that split complex problems arbitrarily, provide explanations to parts and attempt to generalize from the partial observations.

Another split of temporal phenomena is to distinguish time points (time 'stamps', such as now, tomorrow, calendar dates, etc.) from duration (or interval) and use time points for temporal ordering of events (Fraisse 1963; Buhusi and Meck; 2005; Merchant et al. 2013). Notice that such practice also requires to refer to human-made instruments or to world events which themselves are not inherently temporal (such as spring or full moon). A deeper problem is the circular relationship between time as duration and time as an event stamp. Time point is the end or beginning of duration. Conversely, duration is arbitrarily defined by accumulation of units between two time points, analogous to the circular nature of distance and 'here' (i.e., a point is space; Buzsáki and Llinás 2017).

Sequential Brain Activity as a Foundation of Event Ordering

The above discussion of time can bring us back to navigation and to memory, both of which are inherently sequential, and thus may have a relationship to internally generated neuronal assembly sequences (Figure 3.1). There is a clear parallel spatial navigation and episodic memory, which can be referred to as mental or imagined navigation back into the past (Tulving 1972). A main difference is that mental navigation does not depend on immediate environmental or body-reafferent cues. It has been postulated that neuronal mechanisms, which initially depended on external cues in simple organisms, have become 'internalized' and disengaged from locomotion (Buzsáki 2005; Buzsáki and Moser 2013). Without external constraints, disengaged processing in complex brains can create an internalized virtual world and generate new knowledge through vicarious or imagined experience, tested against pre-existing and stored knowledge. This preservation of neuronal sequences can be demonstrated when a rodent is trained in a memory task without the translocation of the head and body. Figure 3.6 illustrates such an experiment in which rats learned to alternate between the left and right arms of a T maze. The key part of the experiment is that a delay is introduced between the choices during which the animal is required to run in a running wheel at a constant speed while facing the same direction on each run (Pastalkova et al. 2008). During running, theta oscillation is sustained and spiking activity of pyramidal neurons display a continuous sequence during the entire journey, with unique sequences for left and right choices. Thus, one can predict the animal's arm choice in the maze tens of seconds before it acts out the plan with high accuracy, including maze arm choices on erroneous trials (Pastalkova et al. 2008). Thus, the unique sequences predict 'what' will happen. The duration, firing rate, theta phase offset relationships between neighboring neurons within the sequence are identical to those of place cells during ambulation in the maze arm.

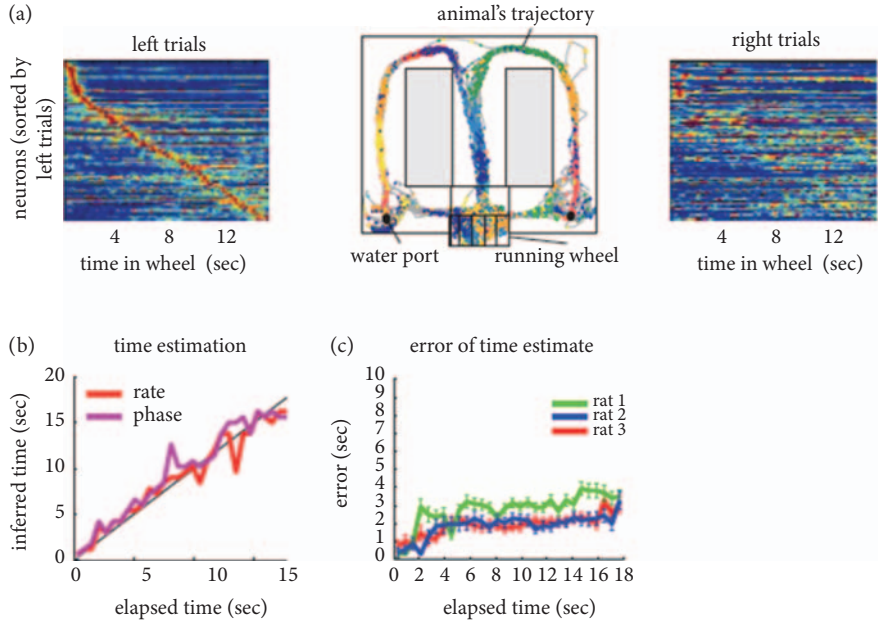


Figure 3.6 Time prediction from sequential neural activity in a memory task. (A) Spontaneous alternation task. The rat was required to make a correct choice for water reward on the basis of its previous choice (middle). Action potentials of different place cells are represented by coloured dots. Left and right, Average raster of firing rates for a population of hippocampal neurons during wheel running. Neurons are ordered according to the time of peak firing rate during left trials. Each line is a neuron, hot colours indicate increased firing rates. (B) Time detection on single trials using a time prediction model fit from all other trials. In each time bin, elapsed time in the running wheel is inferred either from the population firing rate vector (red) or the firing phases of active cells with respect to the theta oscillation (purple). In each case, the prediction approximates well the true time (black). (C) Error of time estimation from population vector of neuronal activity in three rats. Rat 1 had < 50 recorded neurons, rats 2 and 3 > 50 neurons. Note reliable estimation of running duration from neuronal activity. After Itskov et al. (2011).

From the evolving sequences, both in the wheel and the maze, one can also calculate the exact distance travelled by the animal and quantify the run distance against the units of a ruler. Alternatively, one can compare the neuronal sequences against the unit of a clock and report the elapsed time from the beginning of wheel running (Figure 3.6; Itskov et al. 2011). This relationship prompted Eichenbaum to declare that there are 'time cells' in the hippocampus (Eichenbaum 2014). For a moment, it appeared that the long search for the mechanism of episodic memory has been solved by identifying the *what*, *where*, and *when* in the hippocampus (Figure 3.1; Eichenbaum 2014). But this may be only an illusion. Note that in this prototype experiment, there is only one physical change, which is an internally generated sequence. When the same exact neuronal sequence or its individual members are compared with future arm turns, units of rulers and clocks, it appears that they 'represent' different things. But the three names refer to only one mechanism.

Here we need to refer back to the linear relationship between distance, duration and speed. As we have discussed above, knowledge about locomotion speed is critical for advancing neuronal assemblies at appropriate rate of change (Figure 3.5), and neurons, particularly inhibitory interneurons throughout the forebrain, vary their firing rates with speed (Czurko et al. 1999; Kropff et al. 2015). Distances can be computed from the product of the elapsed time and instantaneous speed. Vice versa, time can be computed from product of the distance travelled and instantaneous speed. Why do we expect then that brain circuits 'code' separately speed ('speed cells' Kropff et al. 2015), distance (O'Keefe and Nadel 1978), and time (Eichenbaum 2014), when any of these three variables can be computed from the remaining two? Coding for three separate variables with a linear relationship among them is redundant and would only lead to loss of precision.

Time Is Money: A Social Agreement

We started our overview with the postulate that brains neither sense or make immaterial space or time. Yet, throughout the discussion of the neuroscience observations we used these terms in their everyday meaning, as they refer to units of human-invented instruments. It would be hard to describe any experimental observations without resorting to space and time, the axioms of human thinking. But while physics have made strives to understand the content behind these terms, the vocabulary of neuroscience has remained in the framework of the Newtonian world.

If time is not perceived by the brain why do we feel that time pressure is our number one enemy? Perhaps no other human-invented artifact has had a larger impact on our lives than the clock. Over the past few millennia, time has become the most precious commodity. It all started by measuring distance and duration, that enabled to form a more concrete concept of space and time and their socially agreed metrics. In turn, the ensuing collective knowledge shaped the concept formation of individuals. While the lives of our hunter-gatherer predecessors were filled with timeless ghosts and deities, the invented instruments with their communicable units radically changed thinking of the emerging urban humans. With the help of created artifacts, humans can use their brains in a way that no other animal can approach and using language we can 'fill' the brains of our young with extraordinary amount of semantic knowledge, without the need to go through individual episodic experiences (Buzsáki et al. 2022).

To understand the apparent paradox between reality and feeling of reality, it may be instructive to relate time to another abstract yet more tangible extension of brain function: money. In a sense, money is an abstract representation of value. In its origin, money meant to facilitate exchange of goods when hunter gatherers settled and formed urban communities. Gold coins initially had the equivalent value of exchanged goods but, in today's world, it is simply a symbolic or virtual value based on nothing else but a societal contract. Virtual money, such as a bank

account, is immaterial, and like space and time, cannot exert a direct effect on anything. It has no impact outside civilized humans. A hundred-dollar bill would mean nothing to humans who lived just a few hundred years ago and it has no value even today for isolated societies that live in the Amazon or the Australian desert, not to mention non-human species. Yet, the money concept is regarded as one of the strongest motivational forces, the primary driver of modern societies. But once the social contract for a particular form of currency ends, it becomes irrelevant. Similarly, if human-invented rods and clocks are removed from the arsenal of experimentation, space and time will revert back to dimensionless and ungrounded concepts.

Money has no meaning or utility for an isolated individual because it is not a product of a single brain. This may also be the case for space and time, which terms help communicate our experiences about world events to others. The concepts of space and time, likely money, are the product of a long-evolving societal agreement yet feel very real in our minds. Yet, I assume no neuroscientist would endeavour to study the impact of money per se on the brains of experimental animals. So why are we searching for space and time cells in their brains then? We may not find them. But we will find neuronal assembly sequences, which change at varying rates. Change can occur only to things, such as neurons, and things can exert an effect only on other things. Time is then an abstraction of change, not referring to anything concrete, similar to numbers that are immaterial, abstracted quantities. Instead for searching for mechanisms (*explanans*) of concepts (*explananda*) we created, we may try to revert our thinking by taking the brain as a starting point.

It may well be that self-organized brain networks just do what they can do, which is to generate sequentially active neuronal trajectories, which, in turn, can be matched to our executed (postdiction) and planned (prediction) actions rather than specifying time or space. These same internally generated sequences can index past experiences for recall in the right order or in an imagined or socially agreed coordinate system. The machinery that enabled navigation in the physical world have been exapted for more complex roles, including memory, planning, and imagination, yet navigation in real and mental space appear to use the same general purpose algorithm for tracking sequencies of events and mapping the relationship between continuous variables (Buzsáki and Tingley 2018, 2023).

Actions of animals, simple or complex, are attempts to extract features of our niche and the resulting knowledge is constrained by species specific brain wiring and dynamics. Insects, birds, terrestrial and sea mammals may have very different world concepts, biased by the interactions between their brains and differing niches. Children below the age of 9 intuitively grasp the notion of speed but not time. 'Time relationships constructed by young children are so largely based on what they hear from adults and not on their own experiences' (Piaget 1946). Time is learned through language. The world views of people living across the Andes–Amazonia divide are quite different, as reflected by their distinct Quechua and Arawak languages, respectively (Green 2007). Living in open areas with visible and

identifiable mountain peaks facilitates landmark (allocentric) navigation, whereas if all one can see all the time is just trees in the jungle, egocentric navigation may be more efficient. The different actions taken by people in their own niche may inevitably shape language and communication style with others as well as concept formations about the world. Abstract space and time can afford concrete actions only when we measure them with our instruments using socially agreed units. Only in the context of these cultural innovations can one understand why space and time dominate our current thinking. However, these same innovations place researchers in an difficult situation when they are trying to find the 'representations' of space and time in the brain.

References

Bouffard N., Stokes J., Kramer H. J., & Ekstrom A. D. (2018). Temporal encoding strategies result in boosts to final free recall performance comparable to spatial ones. Mem Cognit. 46(1), 17–31. 10.3758/s13421-017-0742.

Buhusi, C. V., & Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. Nat. Rev. Neurosci., 6, 755–65.

Buzsáki, G. (2005). Theta rhythm of navigation: Link between path integration and landmark navigation, episodic and semantic memory. *Hippocampus*, 15(7), 827–40.

Buzsáki, G. (2019). *The Brain from Inside Out*. Oxford University Press.

Buzsáki G. (2015). Hippocampal sharp wave-ripple: A cognitive biomarker for episodic memory and planning. *Hippocampus*, 25(10), 1073–188.

Buzsáki G. (2010). Neural syntax: Cell assemblies, synapsembles, and readers. *Neuron*, 68(3), 362–85.

Buzsáki G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, 304(5679), 1926–29.

Buzsáki, G., & Mizuseki, K. (2014). The log-dynamic brain: How skewed distributions affect network operations. Nat. Rev. Neurosci., 15(4), 264–78.

Buzsáki G., & Llinás, R. (2017). Space and time in the brain. Science, 358(6362), 482–85.

Buzsáki G., & Voroslakos, M. (2023). Brain rhythms have come of age. Neuron.

Buzsáki G., & Tingley, D. (2023). Cognition from the body-brain partnership: Exaptation of memory. Annu Rev Neurosci. 10.1146/annurev-neuro-101222-110632.

Buzsáki, G., & Tingley, D. (2018). Space and Time: The hippocampus as a sequence generator. Trends Cogn Sci., 22(10), 853–69.

Buzsáki G., & Moser, E. I. (2013). Memory, navigation and theta rhythm in the hippocampal-entorhinal system. Nat. Neurosci., 16(2), 130–38.

Buzsáki, G., McKenzie, S., & Davachi, L. (2022). Neurophysiology of Remembering. Annu Rev Psychol., 73, 187–215.

Czurkó, A., Hirase, H., Csicsvari, J., & Buzsáki, G. (1999). Sustained activation of hippocampal pyramidal cells by 'space clamping' in a running wheel. Eur J Neurosci., 11(1), 344–52.

Das A. (1993). *The Special Theory of Relativity: A Mathematical Exposition*. New York: Springer-Verlag.

Diba K., & Buzsáki, G. (2008). Hippocampal network dynamics constrain the time lag between pyramidal cells across modified environments. J Neurosci, 28(50), 13448–56.

Dragoi, G., & Buzsáki, G. (2006). Temporal encoding of place sequences by hippocampal cell assemblies. *Neuron*, 50(1), 145–57.

Eccles, J. C. (1957). Excitatory and inhibitory synaptic action. Harvey Lect, 51, 1–24.

Eichenbaum, H. (2014). Time cells in the hippocampus: A new dimension for mapping memories. Nat. Rev. Neurosci, 15(11), 732–44.

Einstein A. (1905). On the electrodynamics of moving bodies. *Annalen der Physik* (in German), 10(17), 891–921. 10.1002/andp.19053221004

Ermentrout, G. B., & Kleinfeld, D. (2001). Traveling electrical waves in cortex: Insights from phase dynamics and speculation on a computational role. *Neuron*, 29(1), 33–44.

Fraisse, P. (1963). *The Psychology of Time*. Harper & Row.

Friedman, W. J. (1993). Memory for the time of past events. Psychol. Bull, 113, 44–66.

Galison, P. L. (1979). Minkowski's space-time: From visual thinking to the absolute world. *Historical Studies in the Physical Sciences*, 10, 85–121.

Geisler, C., Robbe, D., Zugaro, M., Sirota, A., & Buzsáki, G. (2007). Hippocampal place cell assemblies are speed-controlled oscillators. Proc Natl Acad Sci U S A, 104(19), 8149–54. 10.1073/pnas.0610121104

Green, S. (2007). Entre lo indio, lo negro, y lo incaico: The spatial hierarchies of difference in multicultural Peru. *The Journal of Latin American and Caribbean Anthropology*, 12, 441–74.

Harris, K. D., Csicsvari, J., Hirase, H., Dragoi, G., & Buzsáki, G. (2003). Organization of cell assemblies in the hippocampus. *Nature*, 424(6948), 552–56.

Helmholtz H. (1850/1948). On the rate of transmission of the nerve impulse. Translated by A. G. Dietze. In W. Dennis (ed.), *Readings in the History of Psychology* (pp. 197–98). New York: Appleton-Century-Crofts. (From Monatsber Akad Wiss Berlin, 21 January 1850.)

Itskov, V., Curto, C., Pastalkova, E., & Buzsáki, G. (2011). Cell assembly sequences arising from spike threshold adaptation keep track of time in the hippocampus. J Neurosci. 31(8), 2828–34.

Johnston, D., & Wu, S. M.-S. (1994). Foundations of Cellular Neurophysiology. The MIT Press.

Kant, E. (1998). *Critique of Pure Reason*, translated by Paul Guyer and Allen Wood. Cambridge: Cambridge University Press.

Kjelstrup, K. B., Solstad, T., Brun, V. H., Hafting, T., Leutgeb, S., Witter, M. P., Moser, E. I., & Moser, M. B. (2008). Finite scale of spatial representation in the hippocampus. *Science*, 321(5885), 140–43.

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience needs behavior: Correcting a reductionist bias. *Neuron*, 93(3), 480–90.

Kropff, E., Carmichael, J. E., Moser, M.B., & Moser, E, I. (2015). Speed cells in the medial entorhinal cortex. *Nature*, 523(7561), 419–24.

Mehonic, A., & Kenyon, A. J. (2022). Brain-inspired computing needs a master plan. *Nature*, 604(1), 255–60. 10.1038/s41586-021-04362.

Merchant, H., Harrington, D. L., & Meck, W. H. (2013). Neural basis of the perception and estimation of time. Annu. Rev. Neurosci., 36, 313–36.

Minkowski, H. (1908). Die Grundgleichungen für die electromagnetischen VorgÄ"ange inbewegten Körpern, Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen (in German), 53–111.

Minkowski, H. (2012). *Space and Time*. Minkowski Institute Press. https://www.minkowskiinstitute.org/mip/MinkowskiFreemiumMIP2012.pdf

Müller, J. (1840/1843). *Elements of Physiology*. 52nd edition, 2 vols., translated by W. Baly. London: Taylor and Walton.

O'Keefe, J., & Nadel, L. (1978). *The Hippocampus as a Cognitive Map*. Oxford, UK: Oxford University Press.

Ornstein, R. E. (1969). *On the Experience of Time*. Penguin.

Petersen, P. C., & Buzsáki, G. (2020). Cooling of medial septum reveals theta phase lag coordination of hippocampal cell assemblies. *Neuron*, 107(4), 731–44.

Pouille, F., & Scanziani, M. (2001). Enforcement of temporal fidelity in pyramidal cells by somatic feed-forward inhibition. *Science*, 293(5532), 1159–63.

Pyenson, L. (1977). Hermann Minkowski and Einstein's special theory of relativity. *Archive for History of Exact Sciences*, 17(1)71–95. 10.1007/BF00348403

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379–423. 10.1002/j.1538-7305.1948.tb01338.x

Shannon, C. E. (1956). The bandwagon. *IRE Transactions in Information Theory*, 2/1(3). http://csc.ucdavis.edu/~cmg/papers/Shannon.IRETransInfoTh1956b.pdf

Sirota, A., Montgomery, S., Fujisawa, S., Isomura, Y., Zugaro, M., & Buzsáki, G. (2008). Entrainment of neocortical neurons and gamma oscillations by the hippocampal theta rhythm. *Neuron*, 60(4), 683–97.

Skaggs, W. E., McNaughton, B. L., Wilson, M. A., & Barnes, C. A. (1996). Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. Hippocampus, 6(2), 149–72.

Somjen, G. (1972). *The Organization of Behavior*. New York: Wiley and Sons. https://www.springer.com/us/book/9781468417074

Strong, S. P., De Ruyter van Steveninck, R. R., Bialek, R., & Koberle, R. (1998). On the application of information theory to neural spike trains. *Pacific Symposium on Biocomputing*, 621–32.

Strong, S. P., Koberle, R., De Ruyter van Steveninck, R. R., & Bialek, W. (1998). Entropy and information in neural spike trains. Phys. Rev. Lett., 80(1), 197–200. 10.1103/PhysRevLett.80.197

Stuart, G., Spruston, N., Sakmann, B., & Häusser, M. (1997). Action potential initiation and backpropagation in neurons of the mammalian CNS. Trends Neurosci, 20(3), 125–31.

Stumpf, C., Petsche, H., & Gogolak, G. (1962). The significance of the rabbit's septum as a relay station between the midbrain and the hippocampus. II. The differential influence of drugs upon both the septal cell firing pattern and the hippocampus theta activity. Electroencephalogr Clin Neurophysiol, 14, 212–19.

Tingley, D., McClain, K., Kaya, E., Carpenter, J., & Buzsáki, G. (2021). A metabolic function of the hippocampal sharp wave-ripple. *Nature*, 597(7874), 82–86.

Tsao, A., Yousefzadeh, S. A., Meck, W. H., Moser, M. B., & Moser, E. I. (2022). The neural bases for timing of durations. Nat Rev Neurosci, 23(11), 646–65.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (eds.), *Organization of Memory* (pp. 381–403). New York: Academic Press.

Végh, J. (2020). Why do we need to introduce temporal behavior in both modern science and modern computing. *Global Journal of Computer Science and Technology: Hardware & Computation,* 20(1), 13–29.

Végh J. (2021). Revising the classic computing paradigm and its technological implementations. *Informatics*, 8(4), 71. 10.3390/informatics8040071

Végh, J., & Berki, A. J. (2021). Why learning and machine learning are different. *Advances in Artificial Intelligence and Machine Learning*, 11(2), 131–48. 10.54364/AAIML.2021.1109

Végh, J., & Berki, A. J. (2022). On the role of speed in technological and biological information transfer for computations. *Acta Biotheoretica*, 70(4), 26. 10.1007/s10441-022-09450-6

Von Neumann, J. (1993). First draft of a report on the EDVAC. *IEEE Annals of the History of Computing*, 15, 427–75. 10.1109/85.238389

4

The Hippocampal Cognitive Map and Episodic Memory

John O'Keefe

Introduction

In this paper I will set out an updated version of the hippocampal cognitive map theory as formulated for the rodent and then speculate as to how it might provide the basis for episodic memory in humans. According to the cognitive map theory, the hippocampus serves 4 distinct but related functions in animals such as the rodent. First, it locates the animal in a familiar environment and relates its location to other locations in that environment to form a map-like representation (Figure 4.1) (O'Keefe and Dostrovsky 1971; O'Keefe and Nadel 1978, 1979). Second, the hippocampus represents the past, present, and potential future contents of each of those locations. Third, it provides a flexible vector-based mechanism for moving towards a desirable location or away from a neutral or undesirable one by any available route. Finally, the map incorporates an internal mechanism for originally creating maps of unknown environments and for updating and incorporating changes to maps of familiar known environments. In the next section, I will briefly elaborate on the first three of these functions and show how they map onto the different activity patterns of the hippocampal pyramidal cells; I will not consider further the 4th, map updating, function. In the second half of the paper, I will propose extensions of the basic spatial mapping system to account for the role of the hippocampus in the storage and retrieval of human episodic memories. Here I am further developing ideas originally presented in (O'Keefe 1993).These will include the addition of (1) a linear time stamp, (2) the ability to shift the current perceived location of the animal in the map to places other than those currently occupied by the perceiver's body, (3) the addition of a representation of the self in the map, and (4) the incorporation of language to the basic spatial memory/navigation system.

The Basic Map Comprises a Collection of Place Representations

As was set out in the original theory(O'Keefe and Nadel 1978, 1979), the primary function of the hippocampus is to provide an interconnected set of

John O'Keefe, *The Hippocampal Cognitive Map and Episodic Memory*. In: *Space, Time, and Memory*. Edited by: Lynn Nadel and Sara Aronowitz, Oxford University Press. © Oxford University Press (2025). DOI: 10.1093/oso/9780192882547.003.0004

place representation. All the different hippocampal functions involve the basic place function of the pyramidal cells in which each location in a familiar environment is represented by the all-or-none firing patterns across large numbers of pyramidal cells in the CA3/CA1 fields of the hippocampus (Figure 4.1). Additional functions are superimposed upon and make use of this basic function. As shown in Figure 4.1, the locations in the map of 1830's Bloomsbury designated by the oval, rectangle, and hexagonal figures would be represented by different patterns of all-or-none activity in the population of pyramidal cells. In small environments, each place cell usually only has a single preferred location but in larger environments, the number of fields increases (Fenton et al. 2008), for example, increasing from 1.3 fields on average in a 0.68-metre cylinder to 4.0 fields in a more complex environment with a 1.4×1.5 m floor. In a 48 m long track 65% of the pyramidal cells have at least one field with many having multiple fields irregularly spaced along the track (Rich et al. 2014). Extrapolating from this suggests that for long 500 m tracks (the range of wild rats) 90% of cells would have one field. In an experiment in which CA3 pyramidal cells were tested in eleven rooms, 210 of 342 (61%) were active in at least one environment (Alme et al. 2014). It seems reasonable to assume that a large number of pyramidal cells have place fields in at least one environment. The same place cell can fire in multiple different environments which means that each location is represented by the pattern of activity across the population of pyramidal cells (Figure 4.1). There appears to be a mechanism for producing coverage of an entire environment within each group of place cells, perhaps using a dynamic first-past-the-post principle where each place cell suppresses its neighbours through inhibitory interneuronal connections ensuring complete coverage of the environment: there are no gaps in the map. In general, most of these pyramidal cells are silent outside the place field but their firing rate in the place field can vary depending on the contents of that location (see next section).

How do place cells identify the relevant locations in an environment and how are they connected together into a single map? They use information provided both by exteroceptive environmental cues and by interoceptive path integration signals. Environmental signals identifying locations probably reach the hippocampus via the entorhinal cortex in the form of boundary-vector, object, and object-vector cells and via the anterior cingulate in the form of simple feature cells. Boundary vector cells respond to the presence of an environmental boundary in a particular allocentric direction from the animal (Burgess et al. 2000; Lever et al. 2009; O'Keefe and Burgess 1996) whereas object vector cells have the same relationship to objects (Deshmukh and Knierim 2013; Hoydal et al. 2019). Boundaries can be constituted by the presence of a large obstruction such as a wall or the edge of an environment such as a sheer drop. Landmark vector cells fire at a particular distance and direction from an object in the environment and do so independently of the identity, size or location of the object or the orientation of the animal's body axis. A small number of neurons in CA1 and CA3 increase firing at the locations where the objects used

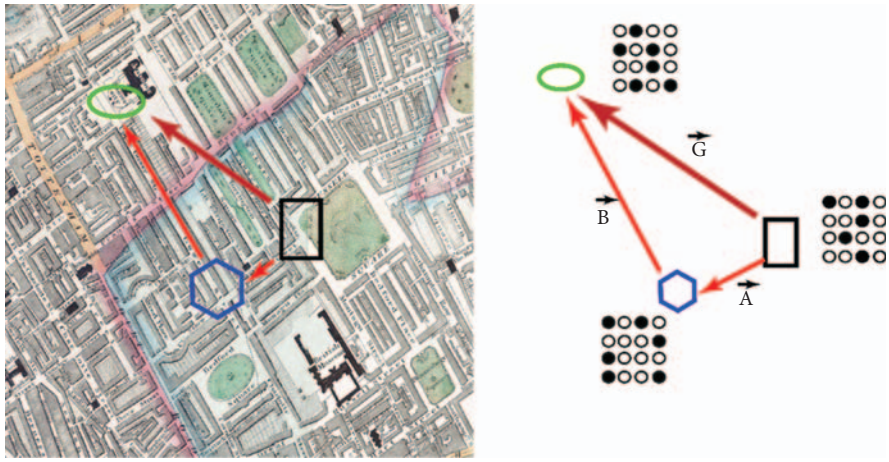


Figure 4.1 Schema of the basic locational and navigational components of the cognitive map theory. Left, superimposed on a map of Bloomsbury London in 1830 are three locations (University College London, green oval; Russell Square, black rectangle; junction of Keppel and Gower Streets, blue hexagon). The crimson arrow shows the direct straight line between the locations in Russell Square and University College and the light red arrows show a possible route between the two locations via the third location at the junction of Keppel and Gower streets. Right, each of the three locations is represented by different activation patterns across a large number of CA1 and CA3 pyramidal cells, here represented by the 4 × 4 set of circles. Black circles are active and white circles are not active. Some active cells are unique to each location, whereas others are active in more than one location (e.g., second cell from the right in the top row). The minimum distance between Russell Square and UCL is represented by the crimson vector G which can be broken down in the mapping system into its component vectors, e.g., A and B to provide alternative routes for getting between RS and UCL. The core of the mapping system is a vector processor capable of decomposing, adding, and subtracting vectors.

to be after they had been removed (Deshmukh and Knierim 2013) presumably providing the basis for hippocampal/subicular object location memories(Poulter et al. 2021).

Place cells firing in an environment are also held together by the fact that they all receive inputs from the same set of head direction cells, maintaining place cells in the same relative orientation to the environment as a consequence (Knierim et al. 1995; Yoganarasimha and Knierim 2005). HD cells (Taube et al. 1990b, 1990a) depend in turn on environmental clues and on vestibular-based path integration inputs and this dependence can be altered by the animal's experience with the relative dependability of these two sources of information (Jeffery and O'Keefe 1999). Movement of a visual cue in the animal's presence can lead to less reliance on that cue and greater reliance on the vestibular system by head direction cells. Finally, when the animal is in one place, movement in a particular direction will tend to activate the place cells representing the neighbouring location. As the animal explores

the environment moving from place to place, the cognitive map builds up a set of transition probabilities between the local place representations perhaps on the basis of place cell by head direction cell activations. I note that recent work questions the role of the head direction system in providing information about direction of movement through the environment (Raudies et al. 2015). Given the evidence of large eye movements compensating for head movements in the free-moving rodent (Meyer et al. 2018; Meyer et al. 2020), it might be useful to look at whether these compensations result in eye-direction rather than head-direction providing a more veridical index of direction-facing and movement direction.

The Cognitive Map Represents the Contents of Places as Well as the Places Themselves

The second important function of the hippocampal map is to represent the contents of the individual locations in the map and this is done by varying the firing rates of some of the cells representing that location (Figure 4.2). These contents can vary from simple features such as the presence of a smell or taste in that location but can be as complicated as the presence of an object, a reward, or a conspecific, see (O'Keefe and Krupic 2021) for a review. When times of occurrence are added to the map (see Figure 4.2), they can also represent events of the past and predictions about events in the future. The prevalence of these feature-in-place responses (for example 31% of place cells responding to odour-in-place following training on an odour-in-place conjunctive task (Komorowski et al. 2009) and 24% of place cells responding to a taste-in-place without any training (Herzog et al. 2019) suggests that many place cells might respond to more than one type of feature in the same place. O'Keefe and Conway (O'Keefe and Conway 1978) showed that different place cells used different combinations of a light, a card, a buzzer, and a fan to identify the place field. There is good evidence that single cue information reaches the hippocampus from the anterior cingulate (AC) (Rajasethupathy et al. 2015). In a recent experiment (Yadav et al. 2022), three sets of cues from four different modalities (auditory, visual, tactile, and olfactory) were presented together forming three different contexts in a virtual environment. Each was accompanied by either a positive, neutral or aversive liquid stimulus. Animals learned to increase licking to the first, inhibit licking to the last, with no change to the neutral stimulus. Each stimulus combination activated hippocampal pyramidal cells presumably because the animal thought they represented three different contexts each with a different valence. Following learning, different cells responded to different combinations of stimuli (e.g., olfactory plus tactile) but almost none (1% v 0.8% chance) to a single stimulus alone. Nearly all neurons that responded to a particular stimulus also showed significant and prominent responses to all other features of the same context group. In contrast, neurons in the anterior cingulate which projects to the hippocampus responded to the single features of the compounds. Inhibition of the AC population silenced the

hippocampal compound cells suggesting that this is the source of the elemental feature input. This finding of responses to particular feature combinations but not to the features themselves lends support to the idea that the hippocampal place cells are part of the solution to the binding problem. The binding problem puzzles over the ability of the nervous system to associate together two or more features of an object with the object when the representation of these features is located in different places in the neocortex (Treisman and Gelade 1980). Faced with a blue cube and a red prism how does one associate the blueness with the cubeness and the redness with the prismness and not vice versa. The cognitive map suggests that the primary association is made on the basis of location so that the representation of the two features in the same place binds them together in the hippocampal representation and secondarily might also serve to bind them together in their original cortical representations.

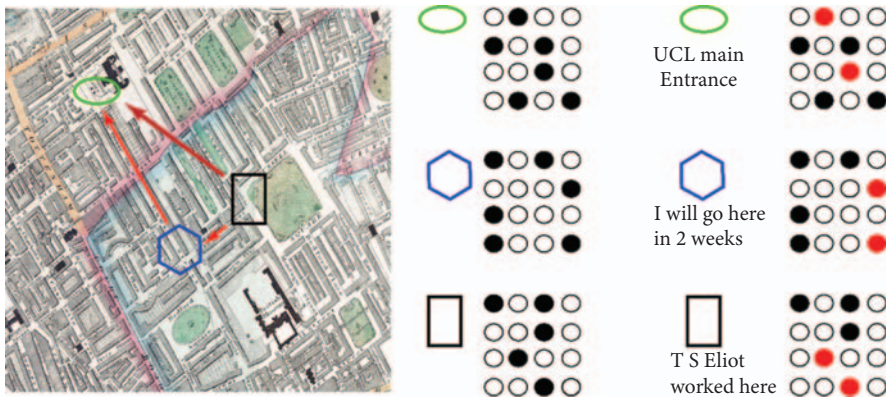


Figure 4.2 Schema of the basic feature-in-locational components of the cognitive map theory. Events occurring in the different locations shown on the left are represented by different firing rate patterns in the place cells. Middle column, as in Figure 4.1: different locations are represented by pattern of place cell firing across population place cells (black circles). Right column: events occurring in these places are represented by increased firings in a subpopulation of the cells representing those places (red circles). Illustrated are (top) feature-in-place: entrance to UCL, green oval; (middle) prediction of a future event, blue hexagon: I will visit this place; and (bottom) episodic memory, black rectangle: I saw a plaque to the poet T. S. Eliot who worked in this place.

Whether rodent place cells can represent the time of occurrence of the feature-location is still uncertain. It is clear that a linear time signal must be incorporated into the hippocampal representation in humans in order to form an episodic memory system but whether this occurs in infra-human animals is still not clear. The original formulation of the cognitive map theory could not see any reason to impute a sense of linear time to non-human animals but evidence since then has suggested that some cognitive capacities of rodents such as those underlying optimal foraging might require this ability. For example, an animal might need to know how long ago

it had visited a feeding site as well as how long it could expect the patch to contain food and this information might be contained in the hippocampal representation of the patch. Evidence in favour of a linear time representation in rodents and birds has been reviewed in (O'Keefe and Krupic 2021). The best evidence that hippocampal pyramidal cells might code for temporal as well as spatial variables comes from single unit studies which suggest that the firing rate within the place field can vary as a function of time in a way similar to firing rate variation with the contents of the place, representing both short time intervals and longer ones. In one set of studies, it was found that place cells increase their activity at a specific time after an event. That is, they might have time fields analogous to place fields and fire after e.g. four seconds have elapsed since a specific 'starting' event such as entering the place field (Kraus et al. 2013; Salz et al. 2016). As far as I know, there is still no evidence that these representations of time are independent of location since the studies usually involve the animal moving in the same place and have not looked at the same time coding in different contexts. An alternative way of coding time might be to represent elapsed time by a systematic deviation from the initial coding for a place or an object-in-a-place either at the level of a single cell or of the population of CA1 cells (Geva et al. 2023; Mankin et al. 2012; Manns et al. 2007; Mau et al. 2018; Ziv et al. 2013). Comparison of this deviated code with the original would act as a measure of time passed. In the Geva experiment (Geva et al. 2023) representation drift in the firing rates was similar in different VR environments sampled at different intervals suggesting that hippocampal place cells signalled the same elapsed time in both environments. In contrast, amount of experience with each environment was signalled by place field location drift.

We can speculate that the information about each specific feature (for example an object or an animal) represented by the feature-in-place cells is broadcast to a large population of pyramidal cells and that the place cells which become feature-in-place cells do so by strengthening the synapses from the feature inputs onto the place cells by some conjunctive process such as LTP. We can see this process in action in the Komorowski (Komorowski et al. 2009) study described above.

Hippocampal Cognitive Map Is the Basis for a Flexible Navigation System

The third function of the hippocampal cognitive map is to provide a flexible navigation system. To do this requires a signal identifying the direction and distance of the goal as well as the animal's current location. Our understanding of this role of the hippocampus in navigation has been clarified by experiments in our laboratory on the honeycomb maze (Wood et al. 2018), a sequential binary-choice dry-land version of the Morris water maze task. The maze consists of thirty-seven or, in its most recent instantiation, sixty-one individually raisable hexagonal platforms. During a typical navigation task, the animal makes its way through a series of binary choices

from different starting locations on the maze to a fixed goal platform. While confined to each platform the animal is presented with two adjacent choice platforms and its optimal performance is to select the platform with the smallest angular deviation from the goal direction. One can think of the task as a sequence of binary choices where the animal's choices are controlled at each choice point in the maze (Figure 4.3).

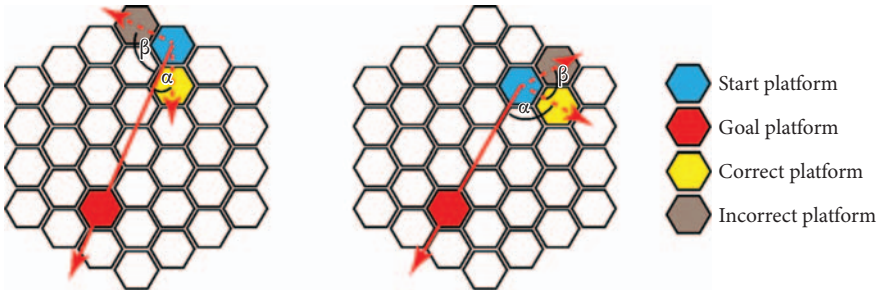


Figure 4.3 Honeycomb maze. Consists of thirty-seven (shown) or sixty-one (not shown) hexagonal platforms configured in an overall hexagonal pattern where each of the platforms can be individually raised or lowered. One platform is designated as a goal (red) and the animal started from different locations on each trial. Each trial consists of a series of binary choices in which the animal is asked to choose between two platforms (yellow and brown) adjacent to the one it is situated on (blue) and is asked to choose the one which is closest to the direction to the goal (in both examples, the yellow one). As in the examples shown, the choice platforms can vary in the angle between them (β) and in the angles relative to the goal direction (α).

Rats with hippocampal lesions are severely deficient in learning to navigate to a goal on the maze and the performance of both control and lesioned animals is adversely affected by several factors: distance of the choice point from the goal, difference in the angle between the two platforms, and deviation of the best choice platforms from the goal heading direction(Wood et al. 2018). On all these measures, hippocampal damaged animals are worse than controls (Figure 4.4). In a follow-up study, we (Ormond and O'Keefe 2022) recorded CA1 place cells on the standard navigation task to see whether they would provide information supporting navigation. Fortuitously, as the animal waits on each platform for the next pair of platforms to arise, it often circles around through a full 360°, enabling us to see whether there is a preferred direction of firing. During navigation on this maze, the firing of CA1 place cells is still localized to a part of the maze but become polarized and is not omnidirectional as is usually found in open field foraging tasks where there is no specific goal; the cells no longer fire equally strongly in all directions in the place field but concentrate their firing in a preferred direction(Figure 4.5A). As the animal moves around within the place field, these preferred directions do not line up in parallel as they would for head-direction cells but change their orientation to remain pointing towards a specific location in the environment. The firing within the place field can be characterized as a vector field pointing within the animal's

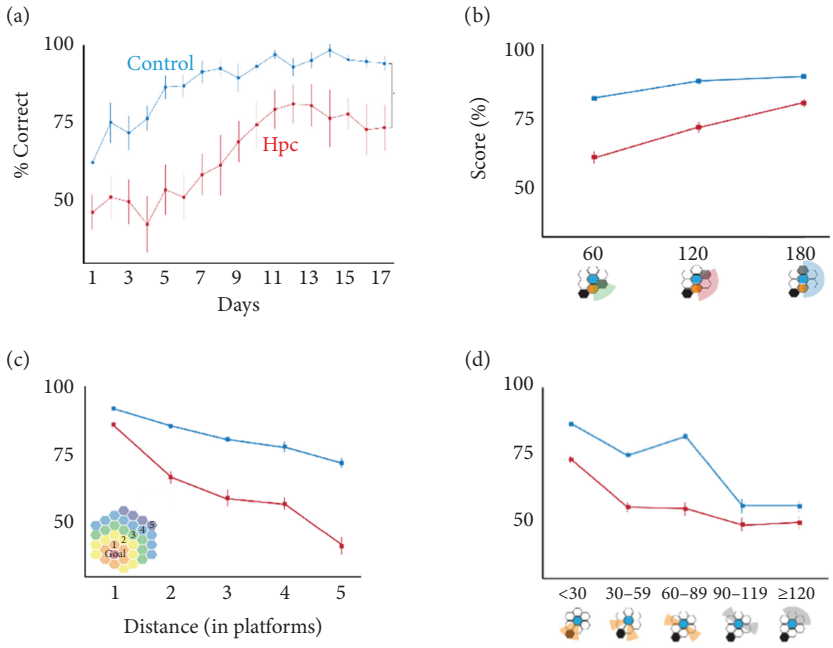


Figure 4.4 Behavioural results on the thirty-seven platform-honeycomb maze. (A) Rats with hippocampal lesions are significantly impaired in learning the navigational task on honeycomb maze. (B) Normal animals (blue) perform worse as the angle between the two choice platforms decreases, and hippocampal-lesioned animals (red) do relatively worse. (C) Normal animals perform worse as the distance from the choice to the goal increases (see inset) and hippocampal-lesioned animals do relatively worse. (D) Normal animals perform worse as the angle between the best choice platform and goal direction increases, and hippocampal-lesioned animals do relatively worse. After Wood et al. (2018).

egocentric polar coordinate framework to a particular location in the environment (called the convergent sink or ConSink) (Figure 4.5B). Surprisingly only a few of these ConSinks are located at the goal itself but most are scattered around the environment at seemingly arbitrary positions(Figure 4.5C). However, importantly, the average of these ConSinks is centred close to the goal location and in front of the animal (0° in egocentric space). Averaging across all of the individual place cell vector fields recorded simultaneously in a given animal creates an average vector field which flows towards the goal from every direction in the environment. Changing the location of the goal changes the location of the ConSinks and their associated vector fields to point to the new goal. Thus the ConSink vector field organization of place cell firing contains sufficient information to solve the honeycomb maze task, i.e., enabling the animal to make correct choices between a platform pointing to the goal and one pointing in a different direction. In an unobstructed environment the animal can reach the goal by following one of these beacon vectors to the goal. It is important to note that the place and ConSink/vector field representations are relatively independent as demonstrated by the fact that in non-polarized environments

in which there is either no goal or many goals there are perfectly good place fields in the absence of strong ConSink representations and vice versa that it is possible to see a large change in ConSink representations together with only a small change in place cell representations after the goal switch in the Ormond and O'Keefe experiments. Importantly, both representations can only be seen at the cell population level and neither the animal's location nor the goal direction can be decoded from the activity of a single cell.

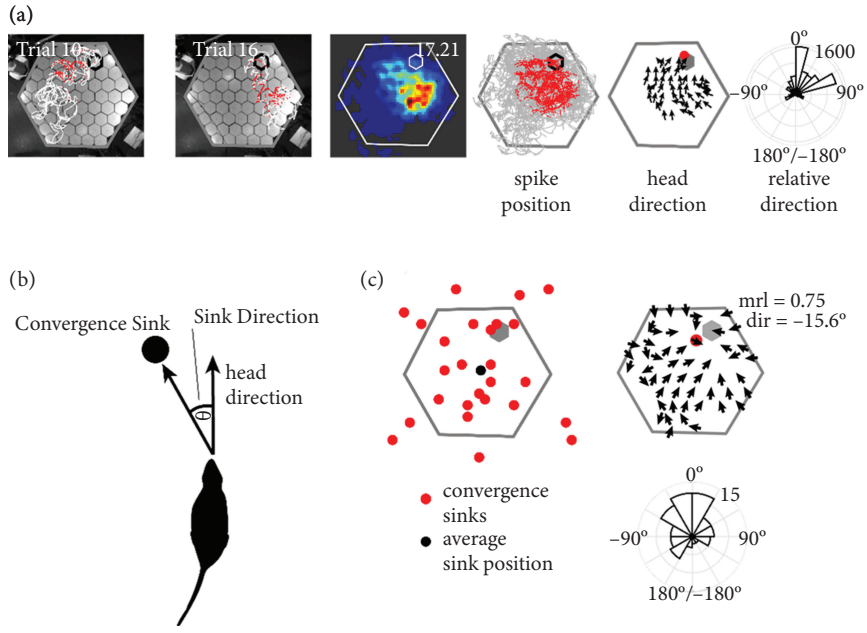


Figure 4.5 Place cell firing during navigation on the honeycomb maze can be characterized by vector fields and convergence sinks (ConSinks). (A) Example Place-ConSink cell showing the tracks of the animal (white) and spikes fired by the cell (red) on two trials as the animal navigates to the goal (black hexagon). The next two panels show the firing field across all trials as either heat map or raw spike and position data. The final two panels show the vector field created by the directional firing at different locations on the maze. The ConSink (red circle) is the point in the environment which best organizes the vector field. The relative direction of the ConSink within egocentric polar coordinate framework centred on the animal's head shows that the best orientation of this ConSink is slightly to the right of the animal's heading direction. (B) is a schematic showing how the ConSink location is organized in the animals egocentric coordinate system for a ConSink which is slightly to the left of centre. All of the ConSinks calculated from the ConSink cells simultaneously recorded in one animal are spread around the environment (red circles, C left). The overall population vector field created by averaging all of the ConSinks from a single animal points towards a ConSink (red circle, C right) located close to the goal (grey circle, C right) and the average sink direction is located in front of the animal (C below). This is typical of all five animals recorded. After Ormond and O'Keefe (2022).

Can the cognitive map provide sufficient information to enable the animal to make correct choices between two platforms neither of which points directly to

the goal or in an environment with obstructions which need to be circumvented, and thus support flexible navigation? A closer look at the vectorial ConSink representation shows that this is the case. As the animal rotates within the place field, the firing typically takes the form of an arc with maximum firing in the centre of the arc, falling off in a monotonic fashion on either side of the maximum (Figure 4.6B).

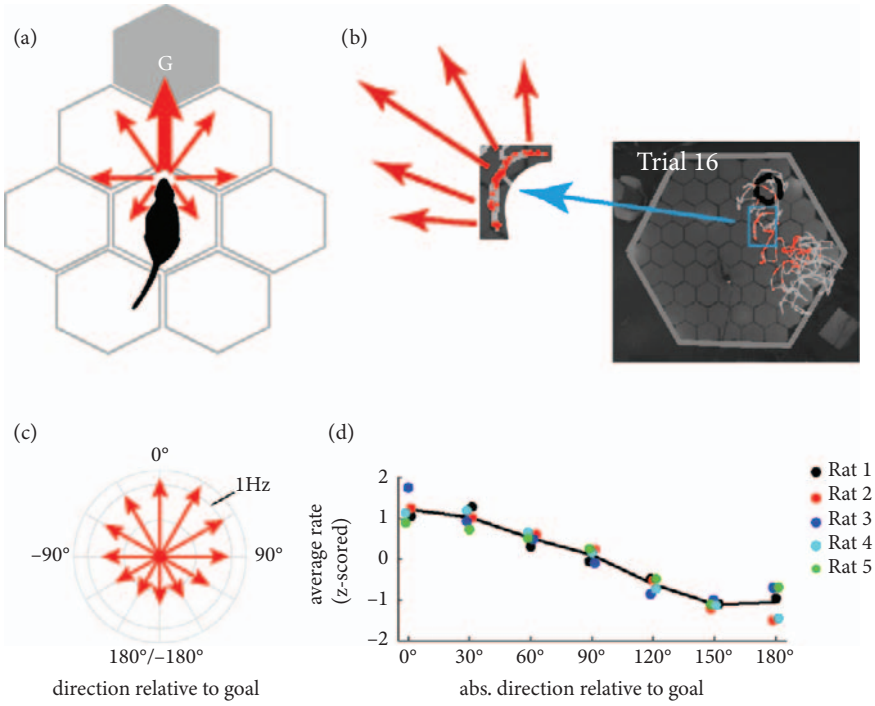


Figure 4.6 The vector field fantail provides a set of vectors pointing in all directions in the animal's egocentric coordinate system which collectively contain information about the straight direction to the goal (A, large red arrow) but also in all other egocentric directions. Importantly, the length of the vectors should fall off in a monotonic function with increasing angle from the goal direction (small red arrows). (B) Fantails are created by the fact that the place cell firing scallops created as an animal rotates around in place on each platform consist of arcs in which the firing increases from a minimum at the beginning of the arc to a maximum when the animal heads towards the ConSink and then diminishes as it continues to rotate, creating a set of vectors whose lengths decrease as a function of angle with the ConSink direction. Summing fantails across five animals results in the figure shown in (C) where there is an orderly falloff with angular deviation from the goal direction. The data from all five animals is highly consistent and creates a sinusoidal pattern. When offered a choice of two or more directions, the system provides information for selecting the optimal direction (in the form of the largest vector) even if that is not directly pointing to the goal. After Ormond and O'Keefe (2022).

This pattern can be characterized as a set of vectors, rather than a single vector, with the amplitude of each vector decreasing with increasing angle from the

ConSink direction(Figure 4.6A). If one averages these vector sets across the population of ConSink cells for a given animal, we find that the largest vector points towards the goal but in addition the average vector lengths decrease monotonically as a function of the increasing angular distance from the goal, forming a fantail configuration(Figure 4.6C, D). Another way of picturing this is shown in Figure 4.7 where the fantail changes orientation to keep it aligned with the goal direction as the animal moves around the environment. Thus, even if the animal is not offered a choice which leads in the direction of the goal, the output of the CA1 field will provide the information about the best choice of the two (or more) offered allowing the animal to navigate efficiently in obstructed environments.

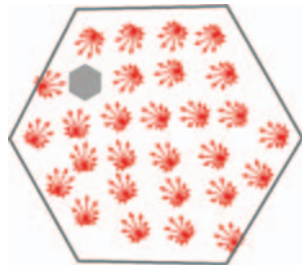


Figure 4.7 Fantail orientation changes as a function of place in a familiar environment providing a valence system about the goal-direction value of all possible directions in all locations. As the animal moves around the environment, the fantail rotates appropriately to maintain the largest vector pointing in the direction of the goal (grey hexagon). Based on data in Ormond and O'Keefe (2022).

Several features of the experiment provide additional information about this fantail representation. First, the experimental paradigm involved a gap of about 10 seconds between the animal arriving at a new platform and it's being afforded a choice of two new choice platforms. During the first four seconds of this period, before the new platforms begin to rise, the animal has no idea which choices it will be offered. During this time, the animal continually rotates in place sampling different directions making it possible to assess the fantail representation prior to the animal having any knowledge about potential choices. On trials in which the animal chooses correctly, the hippocampal fantail representation during this pre-choice period accurately describes the value of the potential alternatives *prior* to information about what choices will be available on that trial (Figure 4.8A left). As the platforms begin to move and the choices become apparent, the fantail representation improves slightly (Figure 4.8B left). It is clear that the fantail firing pattern is an abstract vectorial representation of the valence of possible choice-directions which exist before the actual choices are known. Fantail representations on trials in which the animal subsequently makes an incorrect choice are much less well formed and point on average in a direction other than the goal direction (Figure 4.8A, B right). The absolute firing rates on these error trials are also significantly lower than on correct choice trials (figure 4.8C, D). Although these experiments do not prove that the

animal uses this fantail representation in order to guide its choice behaviour, they are highly suggestive that this is the case.

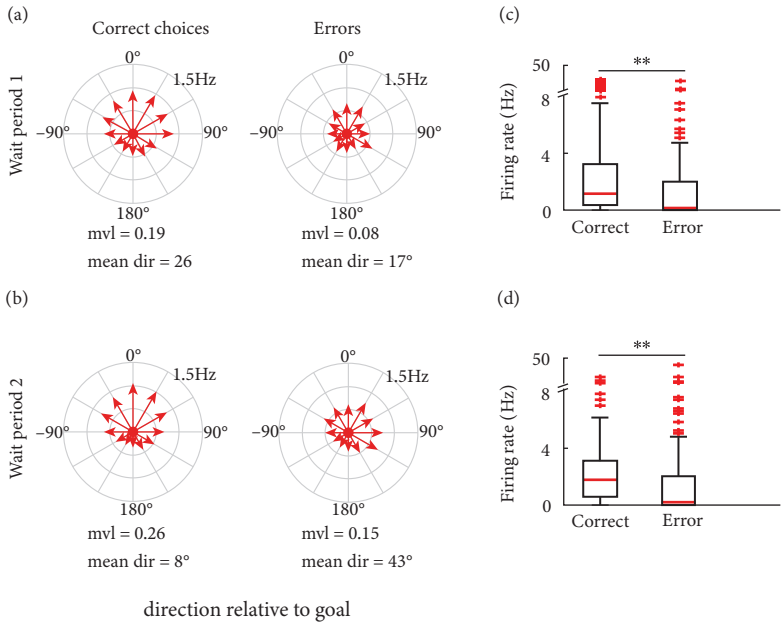


Figure 4.8 Hippocampal fantail is observable on correct choice trials while the animal waits to be given its next choices (A, left). The organization of the fantail improves over the next period as the platform choices become available with mean vector length increasing and angular deviation from the goal decreasing (B, left). In contrast, on trials in which the animal subsequently chooses incorrectly, the fantails are less concentrated (lower mean vector length) and point in a non-goal direction (B, mean direction error 43° in contrast to 8°on correct trials). Firing rates are lower on incorrect trials (C, D). After Ormond and O'Keefe (2022).

Hippocampal Memory Mechanisms

The human hippocampus has long been recognized as a site of memory formation and storage, in particular for memories of an autobiographical or episodic nature. In this section, I will examine the evidence that the rodent hippocampus is the locus of several different types of plastic changes underlying memory. The first type of memory formation is involved in the creation of feature-in-place representations. We have already seen in the Komorowski study described above that during conjunctive odour-in-place training a large number of odour-in-place cells can be recorded following successful learning. During the learning phase of the experiment this number goes from an initial 6% of the cells recorded to 31% when learning is complete. At the same time the total number of simple place cells stays roughly constant throughout training and careful inspection of the data led Komorowski to conclude that these percentages represent the conversion of simple place cells to

odour-in-place cells during learning and the creation of new simple place cells to make up for the loss of the former. Although there are other interpretations, these observations suggest that there are two types of system plasticity learning taking place at the same time over the course of learning. In the first process, inputs to place cells conveying information about the two odours are strengthened. This could be accomplished by something like a simple feature and place associative LTP mechanism. One consequence of this increase in feature-in-place information might be a relative decrease in overall place information and some places in the environment might fail to be adequately represented in the map. Cognitive map theory postulates that there can be no 'holes' in the map representation of an environment and changes need to take place to fill in these newly acquired potential holes. As mentioned above, one mechanism which might accomplish this is a first-past-the-post system in which active place cells representing locations in a particular environment act through the inhibitory interneuron network to suppress other local pyramidal cells from firing in that location. The loss of this pure place cell firing would result in the uncovering of these silent place cells to fill in the potential holes.

Two additional types of plasticity can be seen in the vector field representation during the navigation task when the goal is abruptly shifted, and during successive trajectories to the original goal or after the goal has been shifted. During the goal shift, most of the place cells with vector fields representing the new goal are different from those representing the original goal. Most of the new ConSink representation is carried by previously non-ConSink cells. In the small number of cells which had significant ConSink vector field representations in both goal conditions, eighteen of twenty-eight shifted in the direction of the new goal but ten cells did not (Figure 4.9B). Importantly while the ConSink component of pyramidal cell firing shifts dramatically after goal change, there is only a small (but significant) alteration in place field location suggesting the relative independence of the two processes.

Another important type of plasticity in the ConSink representation is seen within each series of trials to the same goal, whether it be the original goal or the newly shifted goal. Here I focus on the shifted goal data. We did not have enough data to look at changes in ConSink positions on a trial-by-trial basis but instead separated the trials into first and second halves of seven trials each and averaged these to look for any systematic movement (Figure 4.10). On average, there is a large shift of the average ConSink towards the new goal between the first and second halves of the session (Figure 4.10A vs B). This continued movement towards the goal seems to be a general phenomenon. As noted above, movement of ConSinks towards the goal is also seen within the original training to the first goal.

One can think of this as a movement of each ConSink along a vector from its original position towards the goal (Figure 4.10C). As we saw earlier, many of the ConSinks are located off the maze surface and therefore in empty locations which the animal will never have visited. We can speculate that these ConSinks are

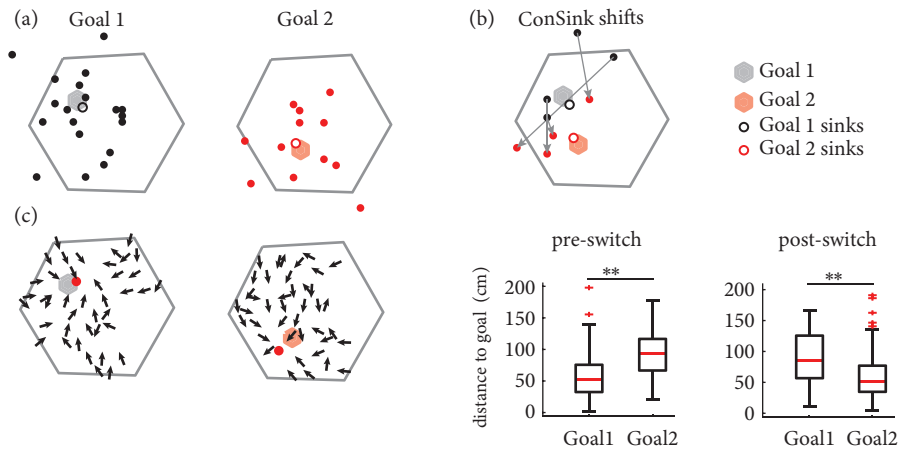


Figure 4.9 Consink/vector field reorients towards the new goal after goal relocation. Example ConSinks (A) and vector fields (C) from one animal showing reorientation to the new Goal 2 (right, pink hexagon) and away from the original Goal 1 (left, grey hexagon) after the animal had been retrained to the new goal. Most of the cells forming the new ConSinks/vector field are different from those involved in the original ConSinks/vector field. The shifts in ConSinks of the small number of ConSink place cells that were involved in both are shown in (B). On average across five animals, the ConSinks before the goal switch are closer to Goal 1 then Goal 2 (D, left), while afterwards they are closer to Goal 2 (D, right). After Ormond and O'Keefe (2022).

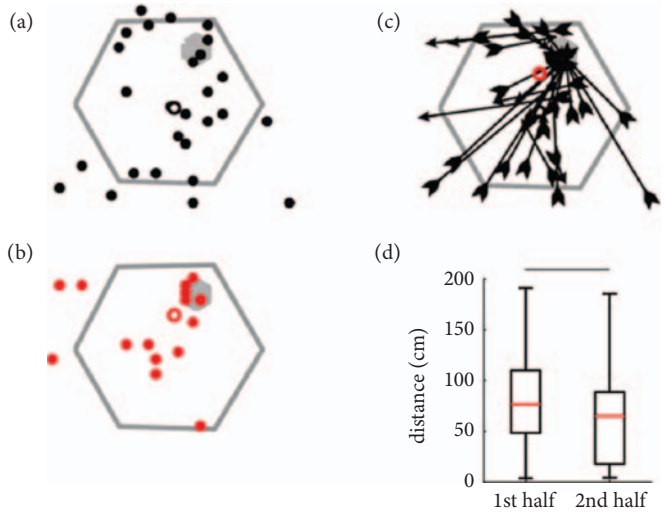


Figure 4.10 Movement of ConSinks closer to Goal with experience. Comparison of ConSinks during the first half of recording session (A) with the second half (B) for a typical animal shows movement of the ConSinks closer to the goal (C). Average of five animals shows significant movement (D). After Ormond and O'Keefe (2022).

determined by environmental cues primarily located distant from the maze. In the present experiment, the use of olfactory and tactile cues on the maze was obviated by the rotation of the maze by + 60° or −60° from its original position between blocks of trials. One way of thinking about the movement of the ConSinks is as an attempt to minimize the difference between the cue configuration as seen from the current ConSink and the same cue configuration as seen from the goal. It is unlikely that this type of sophisticated learning mechanism will be explicable on the basis of simple strengthening or weakening of synaptic connections. For example, one such mechanism might involve the phasor notation in which theta sinusoids represent vectors, vector length relates to sinusoid amplitude, and vector angle relates to phase shift relative to the population theta wave (O'Keefe 1991). On this model, translation and rotation of vectors would involve phase shifts as well as amplitude changes.

Navigation on this view consists of the hippocampus providing the motor system with the current fantail representation and instructing it to choose the path defined by the largest vector available to it. Navigation itself is a sequence of vectors taking the animal from its current location to the goal. The goal is typically defined in the usual animal experiments as ecologically determined primary reinforcers such as food, water or conspecifics but they need not be so. For example, in higher-order conditioning common features such as lights or tones can become goals by secondary association with primary reinforcers. One possible location for the representation of ecological goals is the amygdala which we have shown to contain cells selective for female conspecifics, male conspecifics, foods, or prey-like mobile toys (Mazuski and O'Keefe 2022). This information could come directly from the amygdala or indirectly via the Anterior Cingulate/Prefrontal cortex (Rajasethupathy et al. 2015; Yadav et al. 2022). In the extension to the theory being proposed below, events in the past (episodic memories) can also be identified as goals to be navigated to, either directly or circuitously.

Extension of the Cognitive Map Theory to Humans and Relevance to Human Episodic Memory

From the outset, it was always envisaged that an understanding of the mechanisms of the cognitive map in animals would provide the basis for an understanding of spatial and episodic memory in humans (Burgess et al. 2002). Following Tulving (1983), we define episodic memory as that system which enables us to consciously recollect personally experienced events. The cognitive map would not only store these experiences as instances of visiting particular locations in the past and the events that happened there, but also provide a mechanism for navigating to those representations in a spatiotemporal framework and, as part of this navigation, underpinning some of the psychological properties which Tulving attributed to episodic memory recall such as the concept of the self, autonoetic awareness, and subjectively sensed time, e.g., Tulving (2002). Although Tulving felt that these attributes either did not

exist in animals or, if they did, could not be demonstrated to do so, we shall see that there are strong resonances between the retrieval of episodic memories and some of the mechanisms proposed to underpin navigation in the current paper. In our early writings, Nadel and I proposed that the cognitive map, in addition to its navigational functions, would form the basis for episodic memory in humans with extensions such as the incorporation of a sense of linear time, a representation of the self and the use of the map to store linguistic information as well as information derived from the physical world. We wrote:

> The hippocampus is the neural substrate for one-trial episodic memories in which events are related to each other in a long map extending from the past into the future. These episodic memories can be retrieved directly by the activation of the items themselves, or indirectly by events occurring before or after the event, or by the spatial context in which the event occurred. (O'Keefe and Nadel 1979: 494)
>
> and
>
> The right hippocampus of the human functions in a manner similar to that of the rat, acting as a one-trial episodic memory framework which stores items and events within a spatiotemporal context. The left human hippocampus provides a linguistic framework for the organization of narratives, a framework with properties similar to those which linguists have ascribed to semantic deep structure. (O'Keefe and Nadel 1979: 493)

There is considerable evidence that the human hippocampus is involved in spatial and episodic memory, and navigation (Burgess et al. 2002) and is activated by scene construction, whether real or imagined (Hassabis et al. 2007) and by navigation in virtual reality environments (Maguire et al. 1998; Hassabis et al. 2009; Brown et al. 2010; Hartley et al. 2003). In this section, I want to discuss how our understanding of the role of the hippocampus in navigation might shed light on its role in episodic memory. Let me begin by leveraging our understanding of the role of the hippocampus in flexible navigation in rodents to consider mental spatial travel in humans. In rats, the animal's ego centre is roughly located close to or at the head. I say roughly because there is some evidence that the place cell representation moves a small amount in front of the animal's current location when it forages for food in a two dimensional open field environment (Chaudhuri-Vayalambrone et al. 2023; Muller and Kubie 1989) or when it is trying to make a choice at the junction in a T-maze (Johnson and Redish 2007). In humans it is clearly possible to move one's ego away from the body location as for example in out-of-body experiences. An interesting fMRI decoding study (Guterstam et al. 2015b; Guterstam et al. 2015a) found that the representation of the subject's location in the human hippocampus during out of body experiences sited the subject at his or her subjective location and not at the objective location of their body. Interestingly, episodic memory for events which took place during the out of body experience were less well remembered than those during in body experiences (Bergouignan et al. 2014). In our current understanding of the hippocampal role in flexible navigation, we would say that the centre of the

fantail can be shifted to any location in the environment within the mapping system. Computationally, this mechanism would be similar to the mechanism which is envisaged to guide the animal from its current location to the goal location during navigation. In this case the movement would be a fictive one with no actual movement of the physical body itself. If we now have a mechanism for moving the centre of the fantail to any location in a known environment and also simultaneously moving the representation of the body to the same location, we have the beginnings of a mechanism for episodic memory retrievals.

Of course, episodic memory means retrieving the trace of experiences in the past and for this we need to add a temporal component to the spatial representation. In the original version of the cognitive map theory (O'Keefe and Nadel 1978, 1979), we proposed that the human hippocampus incorporated a linear time stamp which allowed it to store episodic memories as well as purely spatial memories. In contrast we could see no reason why animals such as rodents would need such a time stamp nor was there any evidence to that effect. Subsequently it has become clear that they do possess some temporal component to the spatial system and Julija Krupic and I have summarized that evidence (O'Keefe and Krupic 2021). It is still not clear how this temporal information is incorporated into the cognitive map but the existence of time cells, pyramidal cells which show increased firing after a short period following an event (Kraus et al. 2013; Macdonald et al. 2011; Pastalkova et al. 2008), and the gradual systematic variations in the population firing rates of place cells since the first time of entry to environment suggest possible mechanisms (Mankin et al. 2012; Ziv et al. 2013). We can assume that one of the primary functions of this temporal component to the cognitive map in rodents is to aid the use of the map during foraging behaviour where the animal might need to know in which order it had visited different food patches or how long ago it has been since it visited a particular food patch. Whether this temporal signal rises to the specificity of temporal discrimination necessary for a human episodic memory is not clear. It is still not clear what form this time component needs to take. We have previously discussed the different kinds of temporal representations which have been found in the hippocampal formation of rodents (O'Keefe and Krupic 2021). Introspectively, we humans seem to have a fully metricized temporal representation of events in the last day or two but the metrics quickly deteriorate with time elapsed since our ability to judge the relationship between two events in the past is not much better than representation on an ordinal scale, i.e., one happened before or after the other. Time cells in the hippocampal formation can represent the occurrence of any events up to a few tens of seconds with reasonable temporal fidelity and it is possible that variations in the pattern of activation of place cells over time could be used as a temporal marker but this has not yet been demonstrated. One possibility is that the time signal is treated in the same way as other features and stored as a pattern of firing rates across the place cells representing the location of the past event. Were something like this the case then we could conceive of episodic memory retrieval as

analogous to spatial goal retrieval where the perceiver-based fantails start from the present spatiotemporal location and navigate to find the target memory following a four-dimensional spatiotemporal trajectory. The event in the past would function analogously to the goal in our simple spatial navigation task. Rather than being directed to a physical location, the cognitive map would need to form a fantail starting from the subject's spatial location in the present time and proceed to generate a set of vectors with the largest vector in the direction of the stored spatiotemporal goal. By analogy to the sequence of fantails representing possible pathways to the goal, we would have a fantail sequence directed towards the memory. This would explain the indirect process of retrieval of memory where, for example, one could use the knowledge that the event in question happened near a particular location or some time close to another event in order to 'find' the desired memory. As suggested above that memory would normally consist of the particular event in a particular location at a particular time but could also be associated with a particular person in this case the subject of the experience. Full elaboration of this idea would involve the representation of the subject as being the sort of agent which has a mind and therefore contains spatiotemporal representations of which the desired memory is one such case.

The Cognitive Map and Vector Grammar

One of the great challenges for the cognitive map theory is to show how it can be used for processing and storing linguistic information. Following the original idea in (O'Keefe and Nadel 1978), I have suggested how the map could be used to support a language system called Vector Grammar (O'Keefe 1996, 2003). It starts from the observation that many sentences about physical location and displacement through space make use of prepositions and in the case of inflected languages, the case endings of the nouns, and that most prepositions and many case endings have at their basis spatial, or with a few exceptions temporal, meanings. It follows that if we can show how the cognitive map might represent the ideas of simple spatial location and displacement, we might have the beginning of a computational basis for spatial language. The basic idea is that relationships represented by the prepositions and additionally, the case system of cognitive linguistics (Fillmore 1968) can be captured by sets of vectors, and further that non-spatial uses of prepositions and cases may be achieved by metaphorical extension of the spatial meanings (Lakoff 1980). These vectors were originally thought to be set within an allocentric framework but our more recent work on the use of the rodent cognitive map for navigation suggests that they can also be specified egocentrically. The basic notion is that in addition to storing and manipulating representations identifying objects and their locations as is done in non-human animals, the human cognitive map carries out the same operations on linguistic entities such as nouns.

To take a few simple examples, the deep structure for the sentence:

1) Kristle leaves Russell Square.

would be represented as shown in Figure 4.11 (left) where *Kristle* is represented as moving from that location along an unspecified displacement vector drawn from the divergence vector set represented by the red arrows, and the deep structure for:

2) Kristle goes to University College London

would be represented as shown in Figure 4.11 (right) where *Kristle* is represented as moving to the UCL location along any of the displacement vectors drawn from the convergence vector field represented by the blue arrows. On this reading, *from* specifies the location of the set of tails of the divergent translation vectors and *to* represents the set of heads of the convergent translation vectors. These two vector fields can be represented together as in the sentence about a past event:

3) Kristle left from Russell Square and went to UCL

As can be seen in Figure 4.12, this simple sentence is represented by a large number of potential paths from Russell Square to UCL equivalent to all of the intersections between the two elementary vector fields. Selection amongst these potential paths can be made by the use of the prepositions *via* or *by way of* denoted by the aqua and yellow hexagons. The paths selected by the yellow hexagon corresponds to:

4) Kristle went from Russell Square to University College London via the junction of Keppel and Gower Streets.

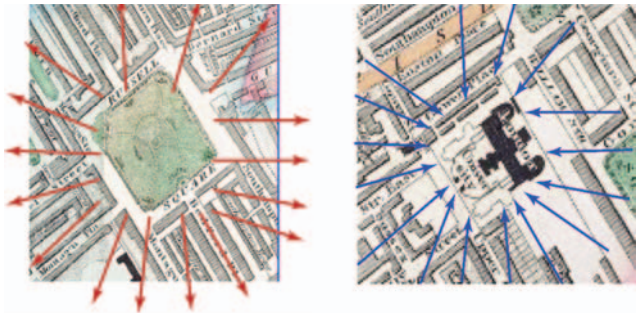
While the path represented by the aqua hexagon corresponds to:

5) Kristle went directly from Russell Square to University College London by way of the shortest possible route.

The use of vector fields rather than simple vectors to represent the meaning of ostensibly simple prepositions such as *from* and *to* serves to capture the ambiguity underlying these prepositions which is one of the hallmarks of language.

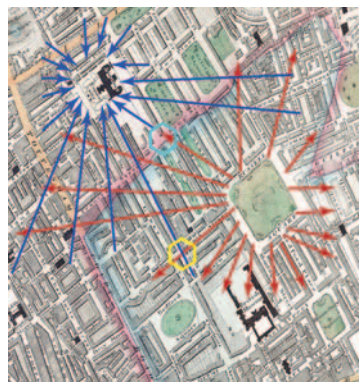
In the examples shown in Figures 4.11 and 4.12, the nominal *Kristle* stands for the item or feature which is translated from one location to the other and in many languages would be represented by the nominative case.

A further elaboration of vector grammar relies on the observation that most of the languages of the world are inflected with different roles specified by different cases, and that most of these cases are either locative, specifying locations of items or translocative specifying the movements of items *from* one location, *to* another location, or *via* a third. In Table 4.1, I have depicted some of the locative cases from the language Tsez, a Northeast Caucasian language with about 15,000 speakers spoken by the Tsez, a Muslim people in the mountainous Tsunta District of southwestern



From Russell Square *To* University College london

Figure 4.11 Prepositions used in descriptions of paths *from* and *to* locations are specified by divergent and convergent vector fields respectively.



From Russell Sq *to* UCL *(via* Keppel St *)*

Figure 4.12 Prepositional phrase used in descriptions of pathways *between* two locations are specified by combinations of divergent and convergent vector fields. The prepositions *via* or *by way of* (yellow and acqua hexagons) further constrain the pathways between the two locations.

Dagestan in Russia and considered to be one of the most complex language existing today. Of the sixty-seven cases in Tsez, fifty-six are spatial. These are formed by combinations of seven locational cases: *in* (in hollow space), *in* (in mass), *under, on, near, near/behind, at/on* (vertical space) (shown in the rows) multiplied by four directional movements relative to those locations: *at, to, from, towards/through* (shown in the columns).

Thus for example, there is a case representing the object *from on top of* which something moves and another representing the object *towards the underside* of which something moves. Furthermore, there are two forms of each of these cases, one of which denotes distal and the other proximal locations. The spatial cases of the nouns in inflected languages can be viewed as reflecting three different aspects of hippocampal function. The *locative* cases (e.g., *in* Russell Square) represent the positional state of the hippocampus reflected in the firing of the place cells when the subject identified by the noun is located in that place. The *ablative* cases (e.g., *away* from Russell Square) represent the (theta-related movement) stage in which there is movement away from the current location in the absence of a specified goal.

Table 4.1 after (Comrie 1998)

Tsez Local Cases (Non-Distal)					
	ESS	**LAT**	**ABL**	**ALLAT**	
IN	-ā	-ā-r	-āy	-ā ɣor	'in (a hollow obj)'
CONT	-ł	-ł-er	-ł-āy	-ł-xor	'in (a mass), among'
SUPER	-λ̄'(o)	-λ̄'o-r	-λ̄'-āy	-λ̄'-āɣor, -λ̄'-ā-r	'on (horizontal)'
SUB	-λ̄	-λ̄-er	-λ̄-āy	-λ̄-xor	'under'
AD	-x(o)	-xo-r	-x-āy	-x-āɣor,-x-ā-r	'at'
APUD	-de	-de-r	-d-āy	-d-āɣor, -d-ā-r	'near'
POSS	-q(o)	-qo-r	-q-āy	-q-āɣor, -q-ā-r	'on (vertical)'
	'at'	'to'	'form'	'towards'	

Tsez Local Cases (Distal)				
	ESS	**LAT**	**ABL**	**ALLAT**
IN	-āz	-āz-a-r	-āz-ay	-āz-a
CONT	-ł-āz	-ł-āz-a-r	-ł-āz-ay	-ł-āz-a
SUPER	-λ̄'-āz	-λ̄'-āz-a-r	-λ̄'-āz-ay	-λ̄'-āz-a
SUB	-λ̄-āz	-λ̄-āz-a-r	-λ̄-āz-ay	-λ̄-āz-a
AD	-x-āz	-x-āz-a-r	-x-āz-ay	-x-āz-a
APUD	-d-āz	-d-āz-a-r	-d-āz-ay	-d-āz-a
POSS	-q-āz	-q-āz-a-r	-q-āz-ay	-q-āz-a

This would be represented by the nondirectional firing of the place cells and can be represented mathematically as the vector field consisting of all vectors with their tails originating at Russell Square as a divergent vector field source (DivSource) (Figure 4.11, left). And finally, the *allative* cases show movement to or into a location (e.g., to or into University College London) (Figure 4.11, right). These cases would be represented mathematically by the vector field consisting of all the vectors with their heads ending at University College or equivalently the convergence vector field sink (ConSink). The combination of the latter two cases (*ablative* and *allative* cases) would identify the vector field consisting of all vectors originating at Russell Square and terminating at University College London (Figure 4.12). The large number of vector sequences specifying this movement (potential paths from Russell Square to University College London) can be greatly reduced by additional information such as a location *via* or *through which* the path must go. One such location (the intersection of Keppel Street and Gower Street) would be specified by another case, the *prelative* case, in a small group of inflected languages but more usually by the prepositions *via* or *through.*

In its simplest form therefore spatial language can be viewed as identifying places, and potential movements to or away from them, with sets of vector fields coding for all the possible vectors originating and terminating in those places together with the identification of intermediate places which would constrain the subset of potentially infinite pathways. Further constraining information could be provided in the form of directional and metrical constraints, e.g., go from Russell Square to UCL by heading north for ½ km.

Summary

The latest version of the cognitive map theory shows how it can be used to identify locations in a familiar environment and to navigate between them in a flexible manner. This involves the generation of vector fields representing the direction and distance to the goal from any location in the environment. These goal-oriented vector fields represent a combination of allocentric (current location) and egocentric (direction of ConSinks) coding which at the neuronal population level provides information about the goal from every location in the environment. This new view of the role of the hippocampus in navigation suggests a mechanism to underpin episodic memory formation and retrieval in humans. The extensions necessary to the rodent hippocampal mapping system to support this role in episodic memory include the ability to code events in the past and to use the navigational capacities of the hippocampus to navigate to those events. An important extension would be the ability to create linguistic events as descriptions of physical events which would enable the human hippocampal system to code for narratives and other linguistic phenomena in addition to locations and navigations in the space of the physical world.

Acknowledgements

My thanks to Mike Hasselmo and Eileen O'Keefe who suggested improvements on an earlier version of this chapter. Research in my laboratory was funded by the Wellcome Trust and the Gatsby Charitable Foundation.

References

Alme, C. B. et al. (2014). Place cells in the hippocampus: Eleven maps for eleven rooms. *Proceedings of the National Academy of Science U S A*, 111(52), 18428–35.

Bergouignan, L., Nyberg, L., and Ehrsson, H. H. (2014). Out-of-body-induced hippocampal amnesia. *Proceedings of the National Academy of Science U S A*, 111(12), 4421–26.

Brown, T. I. et al. (2010). Which way was I going? Contextual retrieval supports the disambiguation of well learned overlapping navigational routes. *Journal of Neuroscience*, 30(21), 7414–22.

Burgess, N., Maguire, E. A., and O'Keefe, J. (2002). The human hippocampus and spatial and episodic memory. *Neuron*, 35(4), 625–41.

Burgess, N. et al. (2000). Predictions derived from modelling the hippocampal role in navigation. *Biological Cybernetics*, 83(1), 301–12.

Chaudhuri-Vayalambrone, P. et al. (2023). Simultaneous representation of multiple time horizons by entorhinal grid cells and CA1 place cells. *Cell Reports*, 42(7), 112716.

Comrie, B., Polinsky, M., and Razabov, R. (1998). *Tsezian Languages*. Harvard University.

Deshmukh, S. S., and Knierim, J. J. (2013). Influence of local objects on hippocampal representations: Landmark vectors and memory. *Hippocampus*, 23(4), 253–67.

Fenton, A. A. et al. (2008). Unmasking the CA1 ensemble place code by exposures to small and large environments: More place cells and multiple, irregularly arranged, and expanded place fields in the larger space 26. *Journal of Neuroscience*, 28(44), 11250–62.

Fillmore, C. F. (1968). The case for case. In R.T.Harms, E.W. Bach(eds.), *Universals in Linguistic Theory* (pp. 1–88). New York: Holt, Reinhart and Winston.

Geva, N. et al. (2023). Time and experience differentially affect distinct aspects of hippocampal representational drift. *Neuron*, 111(15), 2357–66e5.

Guterstam, A. et al. (2015a). Posterior cingulate cortex integrates the senses of self-location and body ownership. *Current Biology*, 25(11), 1416–25.

Guterstam, A. et al. (2015b). Decoding illusory self-location from activity in the human hippocampus. *Frontiers in Human Neuroscience*, 9, 412.

Hartley, T. et al. (2003). The well-worn route and the path less traveled: Distinct neural bases of route following and wayfinding in humans. *Neuron*, 37 (5), 877–88.

Hassabis, D., Kumaran, D., and Maguire, E. A. (2007). Using imagination to understand the neural basis of episodic memory. *Journal of Neuroscience*, 27(52), 14365–74.

Hassabis, D. et al. (2009). Decoding neuronal ensembles in the human hippocampus. *Current Biology*, 19(7), 546–54.

Herzog, L. E. et al. (2019). Interaction of taste and place coding in the hippocampus. *Journal of Neuroscience*, 39(16), 3057–69.

Hoydal, O. A. et al. (2019). Object-vector coding in the medial entorhinal cortex. *Nature*, 568(7752), 400–04.

Jeffery, K. J., and O'Keefe, J. M. (1999). Learned interaction of visual and idiothetic cues in the control of place field orientation. *Experimental Brain Research*, 127(2), 151–61.

Johnson, A., and Redish, A. D. (2007). Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *Journal of Neuroscience*, 27(45), 12176–89.

Knierim, J. J., Kudrimoti, H. S., and McNaughton, B. L. (1995). Place cells, head direction cells, and the learning of landmark stability. *Journal of Neuroscience*, 15(3 Pt 1), 1648–59.

Komorowski, R. W., Manns, J. R., and Eichenbaum, H. (2009). Robust conjunctive item-place coding by hippocampal neurons parallels learning what happens where. *Journal of Neuroscience*, 29(31), 9918–29.

Kraus, B. J. et al. (2013). Hippocampal 'time cells': Time versus path integration. *Neuron*, 78(6), 1090–101.

Lakoff, G., and Johnson M. (1980), *Metaphors We Live By*. Chicago: Chicago University Press.

Lever, C. et al. (2009). Boundary vector cells in the subiculum of the hippocampal formation. *Journal of Neuroscience*, 29(31), 9771–77.

Macdonald, C. J. et al. (2011). Hippocampal 'time cells' bridge the gap in memory for discontiguous events 7, *Neuron*, 71(4), 737–49.

Maguire, E. A. et al. (1998). Knowing where and getting there: a human navigation network. *Science*, 280(5365), 921–24.

Mankin, E. A. et al. (2012). Neuronal code for extended time in the hippocampus. *Proceedings of the National Academy of Science U S A*, 109(47), 19462–67.

Manns, J. R., Howard, M. W., and Eichenbaum, H. (2007). Gradual changes in hippocampal activity support remembering the order of events. *Neuron*, 56(3), 530–40.

Mau, W. et al. (2018). The Same Hippocampal CA1 Population Simultaneously Codes Temporal Information over Multiple Timescales. *Current Biology*, 28(10), 1499–508e4.

Mazuski, C., and O'Keefe, J. (2022). Representation of ethological events by basolateral amygdala neurons. *Cell Reports*, 39(10), 110921.

Meyer, A. F., O'Keefe, J., and Poort, J. (2020). Two distinct types of eye-head coupling in freely moving mice. *Current Biology*, 30(11), 2116–30e6.

Meyer, A. F. et al. (2018). A head-mounted camera system integrates detailed behavioral monitoring with multichannel electrophysiology in freely moving mice. *Neuron*, 100(1), 46–60e7.

Muller, R. U., and Kubie, J. L. (1989). The firing of hippocampal place cells predicts the future position of freely moving rats. *Journal of Neuroscience*, 9(12), 4101–10.

O'Keefe, J. (1991). An allocentric spatial model for the hippocampal cognitive map. *Hippocampus*, 1(3), 230–35.

O'Keefe, J. (1993). Kant and the sea horse: An essay in the neurophilosophy of space. In R. McCarthy, B. Brewer, and N. Eilan (eds.), *Spatial Representation* (pp. 43–64). Cambridge: Blackwells.

O'Keefe, J. (1996). The spatial prepositions in English, vector grammar and the cognitive map theory. In P. Bloom et al. (eds.), *Language and Space* (pp. 277–316). Cambridge, MA: MIT Press.

O'Keefe, J. (2003). Vector grammar, places, and the functional role of the spatial prepositions in English. In E. van der Zee and J. Slack (eds.), *Axes and Vectors in Language and Space* (pp. 69-85) (Oxford: Oxford University Press).

O'Keefe, J., and Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, 34(1), 171–75.

O'Keefe, J., and Nadel, L. (1978). *The Hippocampus as a Cognitive Map*. Oxford University Press.

O'Keefe, J., and Conway, D. H. (1978). Hippocampal place units in the freely moving rat: Why they fire where they fire. *Experimental Brain Research*, 31(4), 573–90.

O'Keefe, J., and Nadel, L. (1979). Precis of O'Keefe and Nadel's *The Hippocampus as a Cognitive Map*. *The Behavioral and Brain Sciences*, 2, 487–533.

O'Keefe, J., and Burgess, N. (1996). Geometric determinants of the place fields of hippocampal neurons. *Nature*, 381(6581), 425–28.

O'Keefe, J., and Krupic, J. (2021). Do hippocampal pyramidal cells respond to nonspatial stimuli?. *Physiological Reviews*, 101(3), 1427–56.

Ormond, J., and O'Keefe, J. (2022). Hippocampal place cells have goal-oriented vector fields during navigation. *Nature*, 607(7920), 741–46.

Pastalkova, E. et al. (2008). Internally generated cell assembly sequences in the rat hippocampus. *Science*, 321(5894), 1322–27.

Poulter, S. et al. (2021). Vector trace cells in the subiculum of the hippocampal formation. *Nature Neuroscience*, 24(2), 266–75.

Rajasethupathy, P. et al. (2015). Projections from neocortex mediate top-down control of memory retrieval. *Nature*, 526(7575), 653–59.

Raudies, F. et al. (2015). Head direction is coded more strongly than movement direction in a population of entorhinal neurons. *Brain Research*, 1621, 355–67.

Rich, P. D., Liaw, H. P., and Lee, A. K. (2014). Place cells. Large environments reveal the statistical structure governing hippocampal representations. *Science*, 345(6198), 814–17.

Salz, D. M. et al. (2016). Time cells in hippocampal area CA3, *Journal of Neuroscience*, 36(28), 7476–84.

Taube, J. S., Muller, R. U., and Ranck, J. B., Jr. (1990a). Head-direction cells recorded from the postsubiculum in freely moving rats. II. Effects of environmental manipulations. *Journal of Neuroscience*, 10(2), 436–47.

Taube, J. S., Muller, R. U., and Ranck, J. B., Jr, (1990b) Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. *Journal of Neuroscience*, 10(2), 420–35.

Treisman, A. M., and Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97–136.

Tulving, E. (1983), *Elements of Episodic Memory*. Oxford: Clarendon Press.

Tulving, E(2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, 53, 1–25.

Wood, R. A. et al. (2018). The honeycomb maze provides a novel test to study hippocampal-dependent spatial navigation. *Nature*, 554(7690), 102–05.

Yadav, N. et al. (2022). Prefrontal feature representations drive memory recall. *Nature*, 608(7921), 153–60.

Yoganarasimha, D., and Knierim, J. J. (2005). Coupling between place cells and head direction cells during relative translations and rotations of distal landmarks. *Experimental Brain Research*, 160(3), 344–59.

Ziv, Y. et al. (2013). Long-term dynamics of CA1 hippocampal place codes. *Nature Neuroscience*, 16(3), 264–66.

5

Space, and Not Time, Provides the Basic Structure of Memory

Sara Aronowitz and Lynn Nadel

Memory makes it possible to apply previously acquired knowledge about the world to one's current circumstances.[1] Having a storehouse of knowledge, however, is but a starting point. Only knowledge that is relevant to the current situation is of use, and one needs a way to pick out what is relevant and what is irrelevant given the quantity of information available. Consider an example: if you were to meet a friend at a familiar art museum, several kinds of knowledge from memory would be relevant. You might want to access memories tied directly or indirectly to your friend, such as the last message she sent you about her job and a story you read online about a political movement that is worrying her. Once at the museum it would be relevant to access information about how the exhibits were organized last time you were there, as well as the fact that the audio guides can be quite helpful. This is a mix between episodic and semantic knowledge, and reflects a wide variety of particular and contingent connections to the present moment. However, among other more context-specific ties, the memory of what happened on previous trips to the museum can be particularly useful.

This example highlights the fact that upon returning to a familiar place, one is highly likely to retrieve memories of what occurred there in the past. Location-cued retrieval is an adaptive approach to determining relevance: to the extent events occurring in the same place follow the same trajectory, memories of the past enable one to anticipate, and act adaptively, in the present. What we will be calling a location-cueing ***heuristic*** attests to a deep linkage between space and memory. As this heuristic is used generally but not universally, we are not proposing that episodic memory is *constitutively* spatial, a view that Phillips (this volume) evaluates and rejects.

This spatial heuristic is obvious almost to the point of triviality. But is there a parallel temporal heuristic? Space and time are often treated together in memory and elsewhere: they are both organizing dimensions of an internal world in which we locate parts of the past, and the cognitive/neural systems underlying spatial and

[1] Thanks to Nora Newcombe and Carl Craver for very helpful comments on a draft, and to all workshop participants for inspiring conversation.

Sara Aronowitz and Lynn Nadel, *Space, and Not Time, Provides the Basic Structure of Memory*. In: *Space, Time, and Memory*. Edited by: Lynn Nadel and Sara Aronowitz, Oxford University Press.
DOI: 10.1093/oso/9780192882547.003.0005

temporal organization may be tightly related or even identical. Yet it would appear that there is no direct temporal parallel for the spatial heuristic we just discussed. This is because one cannot literally return to the same time, in the way that one can return to what one considers to be the same place.

The brain does seem to honour certain kinds of temporal heuristics—one of these is that events happening nearby[2] in time are often related to each other. Temporal contiguity has long been thought to drive associative learning, and the memories resulting from such learning. Neurobiological mechanisms that might partially underlie such temporal contiguity effects have been uncovered, influencing the processes by which neurons are allocated to the encoding of new information (e.g., de Sousa et al., 2021).

A related but distinct temporal structure is order. Temporal order is one ingredient in learning causal and statistical relationships. Order, however, is observed in combination with other event features to build contingency relations that can powerfully drive associative learning (Rescorla & Wagner, 1972). Critically, gauging contingencies demands that organisms aggregate information (statistics) about multiple experiences in an environment, often spread over substantial swaths of time. Doing so, it would seem, requires activating the relevant context representations so that newly acquired information is integrated appropriately.

Temporal contiguity and order are combined in the case of temporal cueing: a temporally contiguous event serves as a reliable cue to a neighbour event in a way that depends on both closeness and order (i.e., the first event is more likely to cue the second event than the reverse). These cues, however, typically share other contextual elements besides time—and when temporally near events are distinct in context, as when they are separated by an event boundary such as switching between two rooms, the cue reliability is weakened (Horner et al., 2016).

These considerations suggest that there is a relationship between time and memory different than the linkage between space and memory. The latter is directly about *where* a remembered event transpired, including information about spatial transformations during the temporal course of the event. The former seems predominantly about these transformations over time, and only secondarily about *when* in an individual's history a given event occurred. Memories of events occur over time, and are about actions that take place sequentially, in a specific temporal order. But they may not be organized around time in the way that memories are organized around space.

In this chapter, we'll start by laying out the problem of relevance in memory in a way that applies to even very basic uses of memory in animals. Within this setting, we'll then explore why and when the location-cueing heuristic is useful. The next

[2] 'Nearby' is of course vague, in that it could be a matter of milliseconds, seconds, hours, or even days. The same issue arises for spatial nearness: in inches, feet, or miles? The relevant scale depends on features of the thinker and the situation, and so there is no absolute notion of nearness in space or in time that captures meaningful nearness across situations.

section will ask: is there a way to think of a temporal heuristic that could capture something about relevance beyond the here and now? In failing to find a version that succeeds, we'll then consider and ultimately dismiss a joint spatial-temporal heuristic. The remainder of the chapter will draw out the consequences of this asymmetry between space and time in memory, and connect this conceptual fact to features of our actual memory systems.

The Relevance Problem

In order to discuss spatial and temporal heuristics in memory, it's first necessary to lay out the problem for which these heuristics are an approximate solution. This is the problem of relevance, of mobilizing information that optimizes one's chance of success in the current situation. Information demands, as just described, may seem reactive—as if demands are always created by the appearance of an unfamiliar situation. But information demands may be anticipatory rather than reactive in nature, such as when you ask yourself how to prepare for what will happen in the museum. In this case, before you enter a given situation, you draw on stored information to make a plan. Thus, information demands are generated throughout a loop between acting, predicting, and observing, rather than following after observation in some absolute sense.

The relevance problem refers to the need to direct memory search and construction in order to satisfy information demands. Consider a rat which encounters a novel object. This object triggers an information demand: has she seen or smelled something similar before? This can be thought of as a sub-question under the demand to know what the object is and what she should do with it, if anything. This demand structures a search through her memories. Since time is limited and her memories are extensive, an answer to the relevance problem addresses how to proceed when seeking to retrieve useful information: where is the best place to start and where should she look next.

An answer to the relevance problem is rarely optimal due to the computational intractability of the problem—we simply have too much stored information to conduct an exhaustive search. Instead, it can be approximated with a heuristic. These heuristics take the form of questions or question-types that are reused across a set of contexts.

What would an optimal solution to the relevance problem look like? The open-ended nature of the information demand makes a clear solution hard to formulate, and in lieu of an optimal solution to the relevance problem, we can see the value of relevance heuristics. They allow us to understand search behaviours in connection to a notion of optimality without supposing a full solution to the relevance problem, and set the stage for our main question: do spatial and temporal relations make for effective relevance heuristics? And are they part of the relevance heuristics that we and other animals actually use to access useful information?

Spatial Heuristics

The critical importance of spatial heuristics is readily seen in the following scenario, that plays out every day in nature: an animal, let's say a common desert rat, is out and about, searching its home territory for something to eat. It enters a region of the territory (call it area A) it hasn't been in for a while. what happens next?

It turns out this is not the only question one should be asking. Another critical question is what happened before? And the answer to that question goes something like this: having felt hungry while nesting, the rat exited the nest with a clear expectation of what it would encounter in its home territory. Using previously stored knowledge about the environment, the rat's brain continuously predicted, in real time, what would happen next. When the rat entered area A, it did not do so with a blank mind, but instead with the expectation that it would find the food it remembers being available from prior experiences in that space, or context.

The question of 'what happened before' is relevant in another way here. In order for the rat to have acquired knowledge about its home territory it had to first learn about it. Let's consider this earlier stage, when the rat, perhaps having been exiled from its former home as a no-longer juvenile, seeks to establish a new home base. It enters many unfamiliar places[3], without specific expectations. But our rat is equipped with some general expectations, about the kind of places that might be safe, or dangerous, for example. Rodents in general are prey species, and hence stay out of the limelight if they can. They explore new places with trepidation, but explore they must, because the acquisition of reliable models of the world—cognitive maps—is essential to survival.

These internal models, built up through exploration of what initially were unfamiliar places, are the basis for the memories that are retrieved when a rat, or a person, re-enters a previously visited place. Without them, one would be maladaptively prone to act in contextually inappropriate ways. With them, one can access relevant information, to answer quite specific questions, only some of which are explicitly about time: what is this place like? Is it a safe place? When was I here last? What happened here before?

All of these questions are centred around places, even as they are also about time. The rat retrieves information related to *this place* from memory in order to generate expectations of what can happen *here*. This happens in two ways. In answer to an information demand, the rat might retrieve a specific event that occurred at that location at a different time, such as the presence of a predator. This retrieval heuristic might be something like 'what happened here last time?' Second, the rat might have formed a general, tenseless representation of this place, for instance, a statistical representation of the likelihood of encountering a predator. This representation, even before it is linked to the current retrieval context, is spatially organized: it combines events at different times in the same space in order to track general features

[3] For the purposes of this chapter, places are assumed to be regions of the environment, not unique points.

of that environment. While distinguishing between event memory and memory for tenseless environmental statistics is important for comparing episodic memory in humans and other animals, among other things, both of these modes of retrieval are inherently spatial.

If spatial retrieval heuristics are common and fundamental, we would expect animals to prioritize both learning about space, and then use spatial information to guide their behaviour in the future. The former is clearly demonstrated by the 'pre-exposure effect' experiments (e.g., Fanselow, 1990), in which rats exposed to a new environment were shocked either shortly after exposure to the environment, or after a two-minute delay. Only the rats who had two minutes to explore and learn about the new context acquired fear of that specific context when shocked in it—the rats who were shocked shortly after exposure to the context did not learn to associate the shock with the specific context in which it occurred, because they were not given sufficient opportunity to first learn about the context. That spatial information is preferentially used to guide action is beautifully shown in a study using hummingbirds, which demonstrated that their behaviour (returning to a previously visited flower) was controlled by where the flower was located, rather than the stimulus characteristics of the flower itself (Hurley and Healy, 1996).

These empirical findings about the importance of space in retrieval heuristics and generalization make sense in light of the learning problem animals face. In most of the natural world, places have substantial inertia. This fact about the natural world did not go unnoticed in the course of evolution—research has shown that it is predominantly the most stable elements in a place that come to identify that place (e.g., Jeffery, 1998). Given that, places either change only gradually or evolve by a function that can be learned. A tree that proves a good source of acorns early in the day will likely still do so later on, and even a month or a year later. This regularity, though common, is not necessary, and some animals live in niches that lack meaningful predictability over the relevant timescales—for instance, Modlinska et al. (2016) observed a colony of rats that lived in a barn feeding on irregularly placed items stored by humans there, and whose nests were often disrupted. These rats, interestingly, did not show the avoidance of new foods (neophobia) that is common in most rodent populations. Similarly, Odling-Smee & Braithwaite (2003) showed that landmarks are only reliable indicators of location for fish in pond-like habitats, which are relatively stable. River-dwelling fish rely much less on landmarks, given that turbulence and river flow render the visual environment unstable. But since most animals live in locations with meaningful regularities, spatial heuristics are generally effective for organizing behaviour. Long-term memory primarily conveys an adaptive advantage through connecting organisms with information beyond the here and now, a kind of action at a distance (cf., Nadel, 2021). Spatial heuristics bridge time to combine information gathered over multiple visits to a place, enabling adaptive behaviour in that place in the future.

To foreshadow our conclusion, spatial and temporal heuristics are different even at this highly theoretical level. Events near each other in space, as we've seen, are

often mutually informative in the environments most animals inhabit. Events near each other in time *but not in space* surely do not share meaningful similarity, nor do events near each other in time across the span of environments that one animal visits from day to day. For a rat who roams over fields, swamps, and forests, there is no reason to suspect events happening in all of these places at midday yesterday have much to do with one another. Agent-independent temporal closeness tends not to be meaningful.

Events occurring near each other in time can be related, such as when the pattern of a gust of wind and a fallen nut might permit an observer to draw a causal inference between the two. But note that these events are not just related in time—they are also both in the same place. This spatial conjunction seems needed for causal connection.

Our felt experience of space is one of staying in the same place from moment to moment, until and unless we cross some physical boundary, such as a doorway (e.g., Radvansky, 2012), at which point we conclude we are in a different place. These spatial boundaries are similar to (and drivers of) 'event' boundaries, but as we see below, there are some crucial differences that have significant implications for what we remember and for how space and time separately interact with memory.

Temporal Heuristics

In contrast to the case with space, where we assume we remain in the same place over time, time never stands still. Our felt experience is one of being inside an 'event' that unfolds over time. In fact, we often use the word 'time' interchangeably with 'event'. When we ask someone 'do you remember the time we went to the museum, or ball game', we don't expect them to retrieve a factual memory of the actual date and time when it happened, though they might be able to do so if asked. Rather, we are prompting them to recall the event, extended in time. Absent some life-changing punctate event, we are generally aware of the steady movement of time within an event rather than having the more molar temporal experience of simply being in that event, with no attention paid to the unfolding specifics. When one is at a baseball game one might sometimes spend one's time thinking about the 'occasion'—a molar description of the event. But often one instead attends primarily to the stuff that is happening in time, moment to moment, then and there. This contrasts with space—here we are aware of where we remain over time, paying little attention to the details that variably define the place from moment to moment.

How does this essential difference influence the ways in which time could plausibly be used by organisms as heuristics? To start, the temporal relations among the elements of an event are often crucial. It matters whether A precedes B or B precedes A—we seem to come equipped with a bias to assume that what comes first can *cause* what comes next, but not vice versa. This is a useful heuristic, and the nervous system takes advantage of it in several ways. For example, a premium is put on learning

sequences, as many (e.g., Dragoi & Buzsaki, 2006; Dragoi, 2020) have shown in the hippocampus and its neighbours, the subiculum and entorhinal cortex. But there are times when temporal order can be misleading. Often events are linked in time of observation just because an organism goes to one place, and then to another, for quite unrelated reasons.

In the next sections, we'll build on the conceptual differences between space and time in retrieval heuristics that we partially spelled out above. Our first task is to consider evidence of how animals actually use time to organize events in memory. As we have seen, a spatial heuristic allows animals to collapse across time and bring together information occurring in the same space. We could imagine several parallel temporal heuristics. In the case of causation already mentioned, I might take a slice of all the events happening just before the effect and generalize across them to specify the causal conditions for the subsequent effect. Is this kind of heuristic commonly applied?

Time in Events: Contiguity and Cyclical Time Indices

While in some cases temporal contiguity facilitates causal attribution (Shanks et al., 1989), this is not universally true. Unsurprisingly, in a paradigm where participants expected the type of effect to take some time to develop after the cause, Buehner & McGregor (2006) found that adults were less likely to attribute causation when the temporal connection between putative cause and effect was tighter. Consider what you would think if you planted a seed in your garden and then saw a fully grown plant five minutes later, and you weren't inside a Jack and the Beanstalk story.

The demise of pure temporal contiguity as the driver of conditioning dates partly to the experiments on 'prepared associations', as best exemplified by the work of Garcia and his colleagues (e.g., Garcia & Koelling, 1966) on what has come to be known as the 'taste aversion' effect. Rats are more likely to associate getting sick with a food (typically novel) ingested some hours ago than with a tone heard a few minutes ago. Similarly, they are more likely to associate foot-shocks received through a grid floor with a tone heard some hours ago than with a food just recently ingested. These classic experiments demonstrated that temporal contiguity can be overridden by other factors, in this case whether the cues and consequences have been linked through phylogeny—establishing heuristics that preferentially associate certain kinds of consequences with certain kinds of stimuli. While heuristics need not hold in every context, these food, tone, and shock relationships are basic to many learning tasks in rodents and, given their centrality in what we know about rodent memory, should not be dismissed as exceptions to a rule.

A very different kind of temporal heuristic would be to link events not in absolute time, but in cyclical measures of time such as a day. Were this a basic memory heuristic, we would expect to see events that occurred in the morning recalled more easily on subsequent mornings rather than nights. Similarly, it would be easier to

learn statistical regularities that depended on the time of day, such as the likelihood of seeing a predator after a birdcall *at dusk*. Such effects have occasionally been reported, within a 'state-dependent' learning framework, but they are small and elusive. Further, while one can imagine a potential environment making such a temporal heuristic adaptive, most animals do not live in such environments. If a generalization of the following sort—birdcall at dawn correlates with food, at dusk correlates with danger—were the rule rather than the exception, animals would profit from employing a heuristic that organized events by time of day. While these types of regularities exist in the environmental niches occupied by some animals, they are not pervasive enough to have elicited, over the course of evolution, widespread learning biases shaped around the day, month, or year.

Time Between Events: Structure, Context, and Order

Another way to think of time as organizing memory is through the temporal relationship between memory encoding and recall. Chater and Brown (2001) argue that memory organization is fundamentally temporal, in that it is structured by the age of a memory. They provide several lines of evidence, but the most relevant to our discussion are: (a) memory errors that involve confusing events that are temporally close, (b) the ability and ease with which animals learn timing-based rules, and (c) rational analysis models including Anderson and Schooler's (1991) 'need probability' account. This type of temporal organization is relative, rather than absolute, and as Chater and Brown note, seems to decrease in significance as memories become more distant.

Do these findings support the conclusion that the memory system is 'chronologically organized'? Anderson's model draws on a statistical relationship between the likelihood that information is needed and the times at which it was last accessed. This parallels, in his example, the relationship between the likelihood that a library book is needed and the times at which it was last accessed. In both cases, a steep forgetting curve makes it such that an item that has not been recalled recently is unlikely to be needed now. But to reflect 'need probability' in a memory system is not necessarily to adopt temporal heuristics in the sense we've been discussing. To see this, imagine a library that uses need probability to modulate retrieval: in this library, the need probability estimate is used to rank books in a search engine, such that lower probability items rank lower. The simplest way to institute such a system would be to weight the estimate of how well the book fits the search criteria by the need probability.

This library would not count as employing a temporal heuristic. Why not? Because the temporal weighting only operates over a separate relevance ranking, making it a modulating factor. By analogy, emotional memories are certainly recalled with higher probability than non-emotional ones, but this does not entail that the structure of memory is emotional. These modulating factors differ in two

key ways from organizing dimensions: (a) they are not used on their own, and (b) they do not meaningfully interact with the other components of the ranking system, in this case the relevance estimate. By contrast, a spatial heuristic does operate alone, for instance, when we have a spontaneous memory of something else that happened in a place, and does meaningfully interact with other elements of memory relevance, such as demarcating places by schematic relationships (the place I'm in now is located spatially in my map of Toronto, but also linked conceptually to other offices, inside other universities, and so on).

This bare notion of time is connected to the notion of temporal context (e.g., Howard & Kahana, 2002). A key feature of temporal context is the idea that, rather than selecting out a particular representation, e.g., linking the cat on the table to the clock striking twelve, some relationships are tied to the entire state of the brain at a time (see De Brigard, this volume; Ranganath, this volume). Unlike every other kind of temporal organization, temporal context is entirely devoid of structure: rather than form a model of when things happen that can be shifted and altered as we learn more, temporal context provides a basic dimension across which brain states are organized. While the temporal context model (TCM) provides useful explanations for such things as serial recall, it remains unclear how, if at all, temporal context can be used to retrieve specific memories.

Temporal heuristics might be employed in the sense that events are organized and remembered with an internal and external order. The event of Ava knocking over the cup and the dog barking is sharply distinguishable from the event of the dog barking and Ava knocking over the cup. Children and adults attend to the ends of events more than the beginnings (Papafragou, 2010; He & Arunachalam, 2017), a regularity which is reflected in many languages. The very concept of an event arguably privileges temporal relationships: events necessarily play out over time, in a specific sequence. And while events also necessarily occur in space, this spatial component can be relatively unimportant.

Does the fundamentality of temporal order in event structure support the idea of a temporal heuristic? Despite appearances, we think it does not. First, the significance of temporal order within events does not imply the same level of salience across events. I might have a library where every book is internally organized by temporal sequence, but where the books are organized by author name (see Hasselmo, this volume).

Second, temporal order within events is neither temporal contiguity (though often involving it), nor any bare measure of location in time: instead, event order is a measure that relates sub-events in a meaningful sequence. For instance, in Papafragou's study, event onsets are glossed as sources, whereas end points are goals: a source (the fairy begins in the tree) and goal (the fairy ends on a flower) are not just organized in time but have a schematic relationship based on assumptions about the causes of the fairy's motion. That is, the end of the motion is a point in time and space that is picked out by its content, not a bare marker—this allows us to adjust it in space and time. Likewise, the event onset picks out a particular part of the event that may have a temporal and spatial address, but is not just an address.

Event boundaries are formed based on information including place and agent-centred factors like surprise. These boundaries, therefore, are not a form of pure temporal organization but a way of organizing events into meaningful units. Thus, we conclude that internal event structure is deeply and essentially temporal, but across-event memory heuristics do not typically draw on temporal order alone.

There is one last form of evidence concerning temporal structure in memory that should be addressed: the well-known retrograde amnesia gradient, in which amnesia seems to effect parts of our access to the past based on time. How could this occur, one might ask, unless the gross structure of memory were in some sense temporal? Recent evidence (see Nadel & Moscovitch, 1997 for an early review) suggests that the retrograde gradient only applies to semantic knowledge—detailed episodic memories are subject to loss in amnesia no matter how remote. But even the existence of a temporal gradient in the impact of amnesia on semantic knowledge does not support the notion that memory is structured temporally in some molar sense. It is entirely unclear how the brain could use the decaying strength of a memory as a useful clue to when something was learned. This would require, at a minimum, that the system was 'aware' of how strong the memory was when first acquired, and beyond that, how much interfering experience had occurred since this initial learning. Neither of these seem biologically plausible.

We've now considered many kinds of temporal involvement in memory, and in each case argued that either the temporal effect is not as widespread as it seems, or that what was presented as a purely temporal relationship is actually equally drawing on non-temporal features such as event structure and spatial context. The fact that these temporal effects are intertwined with non-temporal features does not mean they are insignificant, but it does entail that they are not purely temporal. This matters, as we'll now discuss, because spatial organization can be purely spatial.

Probing the Difference between Space and Time

We have discussed a way in which space, but not time, structures memory. Now we want to consider why this might be. Our hypothesis is that it reflects both conceptual and empirical differences between time and space. By conceptual differences, we mean differences between space and time that would hold for more or less any creature that might exist in a world with the same basic metaphysical structure as ours, rather than purely a priori features. Empirical differences are differences that reflect the spatial and temporal environments in which vertebrates actually developed and currently live. While the task of distinguishing space and time metaphysically is beyond the scope of this paper, we can distinguish space and time conceptually and empirically (see Tables 5.1, 5.2).

Repeatability is about the uniqueness of the spatial and temporal index. Spatial indices are not unique, meaning that just by telling you where something happened to me, I haven't necessarily picked out a unique event since many things can happen

Table 5.1 Conceptual differences

Feature	Space	Time
Repeatable	Yes	No
Direction of motion	Any (no privileged direction)	One direction
Control of motion	Often	Never

Table 5.2 Empirical differences

Feature	Space	Time
Meaningful statistical correlations	Often	Rarely
Presence of landmarks	Almost always	Rarely
Flow/passage	Marked	Inferred

to me in the same place. Time lacks repeatability because, while many events unfold simultaneously, we only have one experience at each time. The direction of motion refers to the fact that there is no canonical direction that animals move in space, whereas we all move forward in time. Time is deeply asymmetrical with respect to our experience. Relatedly, control of motion means that while we often, but not always, decide how to move in space, our motion and speed in time is not up to us. All of these differences may depend on our contingent natures, but it's hard to see how they could be otherwise.

On the other hand, the empirical differences between space and time above are easier to imagine changing. Space is a site of useful statistical correlations in many environments, such as when a mountaintop area has a different distribution of animal activity and plant species than other locations. A fish living in a fish tank would find much to rely on in terms of correlations between parts of the tank and things that matter to a fish such as food and safety. Temporal regions are not usually so informative for us, but there are certainly environments that are very different at different points in cyclical time, such as a tidepool. Passing through space, in most human environments, means passing through visible, meaningful landmarks, whereas the passage of time is measured relative not to landmarks but to intrinsic temporal features like calendar time. However, there are some cases of temporal landmarks, such as sunset, and some cases where space is unmarked, such as an aerial environment with poor visual acuity. Note, however, that while we can think of cases of temporal landmarks in cyclical time, it's hard to imagine objective time having a perceptible landmark, e.g., a visual or sensory feature that lets us know it must be this particular time. The final feature we list is closely related to landmarks: the feeling of the passage of space is given by, for instance, optical flow (Lee, 1980), whereas the passage of time is thought to be tracked internally or inferred indirectly (Ivry & Sclerf, 2008).

We highlight these two kinds of differences in connection to our main argument. Many of the differences between the role of space and time in retrieval heuristics reflect at least one of these features. For instance, Brown and Chater utilize repeatability in their argument for the fundamentality of time, noting that a key difference between space and time is that events have a 'unique address' in egocentric time, whereas spatial locations are repeatable features of objects. While they view this as a reason why temporal organization should be more fundamental than spatial organization, for retrieval heuristics and statistical learning, the opposite would appear to be true. A unique address in egocentric time does not allow events to be brought together, whereas the spatial memory mechanisms we've been discussing can bring a set of events together based on their shared location.

These empirical differences between space and time can be linked to memory differences as well. The difference in generalizations tied to space versus time explains why the spatial retrieval heuristic is useful and the temporal one less so: by bringing together past encounters at different times in the same space, the spatial heuristic allows us to learn and use these generalizations. The presence of external perceptual cues for landmarks and flow in space means that this process can work accurately, since one memory of a space can be linked to another of the same space by taking advantage of these similarities. Our ability to re-identify places is dependent on these capacities to identify space in the first place. When it comes to time, on the other hand, it is hard to take two detailed event descriptions and determine that they were simultaneous in time or separated in time unless the event descriptions contained explicit temporal information or enough indirect details to make a guess about order (such as a person having long hair in one event and short hair in the other).

For animal-like creatures in any environment, space and time have some differences: time moves only in one direction, does not repeat, and its passage is outside the control of the creature. These are compounded by the differences between space and time in our actual environments, where spatial associations are both more significant and more learnable than temporal associations due to the presence of landmarks and signs of the passage of space, along with meaningful correlations. These two types of comparisons help make sense of the differences we've observed in memory retrieval: it is no surprise that we tie generalizations to space and use location-cueing as a frequent default, since space is a useful structure for linking events.

What Does This Tell Us About Memory Organization?

In contexts ranging from theoretical physics to organizing a library book collection, space and time are dimensions that jointly structure a four-dimensional space. Here space and time play the same structural role, and perhaps even are the same in a deeper sense. But as we've shown, the organizational roles of space and time

in memory retrieval are not like this. Instead, a default and central organizational feature of memory links events that occur in the same place across different times. This fact might explain the overlap between the neural systems engaged in learning about or navigating through spaces, and our memory for events—the very experiences that generate the internal models required to efficiently move through the world.

With the exception of memory for events, our encoding of time in memory is typically implicit—action sequences, for example, are precisely encoded and retrieved, but we are unaware of these underlying mechanics and cannot use them to retrieve a sense of time. The passage of time in a particular space does not change our understanding of that space—when we retrieve a memory of a place that we've visited, our sense of that space is not noticeably altered by how much time we've spent in it. In contrast, our estimation of both elapsed time and the temporal order of events is strongly influenced by changes in context, typically spatial location (Horner et al., 2016).

Another way in which time is less stable than space is across the dimension of value. We discount the value of expected rewards the further out in time they may be expected to occur, but the value of a given place remains the same, independent of time, unless some very significant change happens. Paris is, and in one's experience typically remains, Paris. This fact reflects a point we made earlier—it is the most stable elements that define a place.

This leads us to the key point. In all the ways we have described, placing an event in space is already to begin to generalize. This is not true of placing an event in time. As we've detailed, putting an event in spatial context is the basis of learning regularities of that particular environment. Making links between events across time by their spatial correspondence is thus a form of abstracting away from time.

If a rat encounters a tasty piece of pizza in a subway station several times across a few visits, she first co-locates these separate events in space. But once she has, and starts to think of the place itself as rewarding, she is now modelling a feature of that place independently from a particular trajectory she took in the environment on a particular visit. In this sense, her representation goes from one that is centred on herself, to one that is centred around a feature of the environment. Situating an event in time does not have this feature, in part due to the nature of time as we experience it (as not repeatable), and in part due to the nature of the things we're most often interested in learning (properties of places rather than properties of cyclical time alone). In line with O'Keefe and Nadel (1978, 1979), this explanation takes for granted an allocentric understanding of space.

Seeing spatial location as a basic form of generalization has consequences for the episodic/semantic distinction, for the role of the hippocampus in schematic knowledge, and for the relationship between human memory and that of other animals. We'll discuss each in turn.

First, seeing spatial organization as fundamental in memory and also a form of generalization emphasizes a functional continuity between episodic and

semantic memory. Tulving's discussion of the features of episodic memory includes that they reference personally experienced episodes (2002). Our conclusion, on the other hand, suggests that when events are organized in space along with time, they are related to other events in that same space in ways that are neither unique nor essentially reference personal experience. When we situate an event in space and time, as Tulving would have it in the case of episodic memory, we also situate it in space alone, and by doing so build a model in memory that is not solely episodic. Computational approaches (Kumaran & McClelland, 2012) and empirical evidence (e.g., Schapiro et al., 2016, 2017) show that the hippocampal system, central to episodic memory, is also capable of deriving statistical regularities across multiple related experiences. Thus by dividing spatial episodic memory from general semantic memory, we are missing a conceptual and cognitive connection between space and generalization.

Finally, we can now deal with a puzzling loose end. We've argued that spatial heuristics are a basic and fundamental form of memory retrieval, and our evidence for this mostly comes from work on non-human animals. In an adult human, the question of what comes to mind by default is almost impossible to answer. Seeing a piece of paper float over train tracks might remind you immediately and directly of a documentary you saw seven years ago about an anthropomorphized 'morning wind' that blows around Iran. Humans, that is, often engage with our environments in ways that are mediated by a particular and complex set of abstract concepts. The concept of a morning wind is a temporal one, and it is also highly abstract—for instance, in the case of the Iranian documentary, the morning wind is linked to myths about nature. It is hard to imagine similar concepts in other animals, though perhaps not impossible.

This explains why our conclusion, that temporal location does not necessarily involve generalization the way that spatial location does, is not blatantly obvious. When we locate events in time, we can also link them in complex ways to temporal concepts, and doing so is of course a kind of generalization. We can learn regularities such as the kind of thoughts one has as a teenager, the way a friend acts before he drinks his coffee, and so on. But it would be a mistake to assume that these ways of thinking about time provide meaningful structure to memory. As we've shown, there is a more basic form of temporal location, distinct from generalization, that does not serve as a default memory heuristic or structure. The differences played by space and time in organizing memory encoding and retrieval are revealed at this basic level.

We have considered behavioural and neural evidence in animals to argue that spatial relationships are treated as default heuristics in retrieval and structure learning in ways that temporal relationships are not. The history of the study of animal learning demonstrates the inability of temporal contiguity to serve as a core learning dimension—while important in many instances, temporal contiguity alone cannot provide a useful heuristic in structuring memory. We conclude that space plays a fundamental role in memory processes that is not matched by time. It provides the

basis for important generalizations that facilitate adaptive updating of our knowledge base, and it does this by providing a critical heuristic that allows for efficient search and retrieval through this base. Space, and not time, is the backbone of memory search and the beginning of generalization.

References

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, 2(6), 396–408.

Buehner, M. J., & McGregor, S. (2006). Temporal delays can facilitate causal attribution: Towards a general timeframe bias in causal induction. *Thinking & Reasoning*, 12(4), 353–78.

Brown, G. D., & Chater, N. (2001). *The chronological organization of memory*: Common psychological foundations for remembering and timing.

de Sousa, A. F., Chowdhury, A., & Silva, A. J. (2021). Dimensions and mechanisms of memory organization. *Neuron*, 109(17), 2649–62.

Dragoi, G. (2020). Cell assemblies, sequences and temporal coding in the hippocampus. *Current Opinion in Neurobiology*, 64, 111–18.

Dragoi, G., & Buzsáki, G. (2006). Temporal encoding of place sequences by hippocampal cell assemblies. *Neuron*, 50(1), 145–57.

Fanselow, M. S. (1990). Factors governing one-trial contextual conditioning. *Animal Learning & Behavior*, 18, 264–70.

Garcia, J., & Koelling, R. A. (1966). Relation of cue to consequence in avoidance learning. *Psychonomic Science*, 4, 123–24.

He, A. X., & Arunachalam, S. (2017). Word learning mechanisms. *Wiley Interdisciplinary Reviews: Cognitive Science*, 8(4), e1435.

Horner, A. J., Bisby, J. A., Wang, A., Bogus, K., & Burgess, N. (2016). The role of spatial boundaries in shaping long-term event representations. *Cognition*, 154, 151–64.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, 46(3), 269–99.

Hurly, A. T., & Healy, S. D. (1996). Memory for flowers in rufous hummingbirds: Location or local visual cues? *Animal Behaviour*, 51(5), 1149–57.

Ivry, R. B., & Schlerf, J. E. (2008). Dedicated and intrinsic models of time perception. *Trends in Cognitive Sciences*, 12(7), 273–80.

Jeffery, K. J. (1998). Learning of landmark stability and instability by hippocampal place cells. *Neuropharmacology*, 37(4–5), 677–87.

Kumaran, D., & McClelland, J. L. (2012). Generalization through the recurrent interaction of episodic memories: A model of the hippocampal system. *Psychological Rreview*, 119, 573–616.

Lee, D. N. (1980). The optic flow field: The foundation of vision. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 290(1038), 169–79.

Modlinska, K., & Stryjek, R. (2016). Food neophobia in wild rats (Rattus norvegicus) inhabiting a changeable environment—A field study. *PloS one*, 11(6), e0156741. https://doi.org/10.1371/journal.pone.0156741

Nadel, L. (2021). The hippocampal formation and action at a distance. *Proceedings of the National Academy of Sciences*, 118(51), e2119670118.

Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current opinion in neurobiology*, 7(2), 217–27.

Odling-Smee, L., & Braithwaite, V. A. (2003). The influence of habitat stability on landmark use during spatial learning in three-spined stickleback. *Animal Behaviour*, 65, 701–07.

O'Keefe, J. and Nadel, L. (1978) The hippocampus as a cognitive map. Oxford: The Clarendon Press.

O'Keefe, J. and Nadel, L. (1979) Precis of O'Keefe and Nadel's The hippocampus as a cognitive map, and Author's Response to Commentaries. *The Behavioral and Brain Sciences* 2: 487-534.

Papafragou, A. (2010). Source-goal asymmetries in motion representation: Implications for language production and comprehension. *Cognitive Science*, 34(6), 1064–92.

Radvansky, G. A. (2012). Across the event horizon. *Current Directions in Psychological Science*, 21(4), 269–72.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokosy (Eds.), Classical conditioning II: Current research and theory (pp. 64–99). New York: Appleton-Century-Crofts.

Schapiro, A. C., Turk-Browne, N. B., Norman, K. A., & Botvinick, M. M. (2016). Statistical learning of temporal community structure in the hippocampus. *Hippocampus*, 26, 3–8.

Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: A neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372, 20160049.

Shanks, D. R., Pearson, S. M., & Dickinson, A. (1989). Temporal contiguity and the judgement of causality by human subjects. *The Quarterly Journal of Experimental Psychology*, 41(2), 139–59.

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

PART II

6

A Place for the Memory Trace

Sarah Robins

Introduction

Aristotle characterized memory traces as impressions in wax.[1] Experiences are stamped on the mind 'just as persons do who make an impression with a seal' (*de Memoria et Reminiscentia,* 450a). The successful formation, retention, and reanimation of such a trace requires that the mind's wax be in the right condition: just malleable enough. Too firm and it will fail to leave behind any impression at all. Too soft and it will be easily overridden by subsequent impressions.

Aristotle's depiction is, on the one hand, quaint. On the other, it shares important similarities with contemporary theorizing about memory. When pressed to define memory, appeals to representations remain a common feature. Take, for example, Yadin Dudai's (2007) definition of memory as 'the retention over time of experience-dependent representations' (p. 15), or Morris Moscovitch's (2007) construal of memory as the 'lasting internal representation of a past event or experience (or some aspect of it) that is reflected in thought or behavior' (p. 17).

Wax of course no longer works as a metaphor. Wax tablets were traded in for parchment and then for paper. Subsequent upgrades include the camera obscura, the phonographic record, the computer hard drive, the cloud.[2] While the metaphors have changed, the reliance on metaphor to build an analogy between memory and forms of information encoding, recording, and transfer continues. The memory trace thus remains as puzzling as it is persistent.

The lack of attention to the memory trace concept is shared amongst philosophers, psychologists, and neuroscientists—although the oversight manifests differently across disciplines. Philosophers are increasingly wary of memory traces and the need for them to account for memory. This is perhaps best evidenced by the recent emergence and growing popularity of *simulationist* accounts of memory, according to which memory is a way of imagining the past—a capacity whose reliability is not underwritten by traces (e.g., Michaelian 2021; Munro 2021). Even those who want to keep memory traces make a point of diluting the concept so as to render it more palatable. For example, some propose that traces are dispositional, rather

[1] Plato first proposed, but dismissed, the wax tablet view in the *Theatetus.*
[2] See Draaisma (2000) for an excellent overview of memory metaphors throughout history.

Sarah Robins, *A Place for the Memory Trace*. In: *Space, Time, and Memory*. Edited by: Lynn Nadel and Sara Aronowitz, Oxford University Press. DOI: 10.1093/oso/9780192882547.003.0006

than explicit (De Brigard 2020, this volume). Others pull back from the idea of localization, characterizing traces as distributed (Bernecker 2010). Still others cast the scope of traces wider, eschewing their dependence on the mind/brain and allowing traces to be embodied and extended (Sutton 2008). And still others keep the focus on traces as internal, but construe them as content free (Hutto & Peeters 2018) or minimal (Werning 2020). The range of proposals on offer reveals an underlying issue. While many who continue to think of memory traces as important, it is unclear whether there is any further consensus amongst them about what traces are, or the features of them that are critical to retain. The lack of internal agreement raises further questions about what is being rejected by simulationism and other purportedly non-trace accounts and whether the disagreements are substantive or semantic.

Amongst psychologists and neuroscientists, the commitment to memory traces remains more firmly in place, but there is little explicit attention given to what this entails. This is due in part to interest in exploring memory's role in a broader understanding of cognitive and neural dynamics—e.g., how memory is influenced by the brain's oscillatory dynamics in general and the production of spatiotemporal trajectories in particular (Buzsaki 2019), or accounts of episodic memory as a byproduct of a larger system for episodic simulation (Schacter & Addis 2007) or scene construction (Maguire & Mullally 2013). The neglect is also due to the concept's centrality; it is often simply 'baked in' to how memory is understood. Consider, for example, the book *Science of Memory: Concepts* (Roediger, Dudai, & Fitzpatrick 2007), which identifies sixteen key concepts for memory science and features short reflections from three or four leading researchers on that concept. The list of concepts includes encoding, transfer, forgetting … but not the memory trace. Memory traces do, however, find their way into the book. They are the focus of entries on the introductory concept: memory. The quotations from Dudai and Moscovitch at the outset of this chapter were, in fact, drawn from those entries. This is of course only an anecdote, but it illustrates how the importance of a concept can coexist with its neglect. Memory, the capacity, involves retention of memories. A commitment at this level of generality is sufficient for extensive and fruitful investigation of the factors that influence what and whether remembering occurs and identification of the underlying cognitive and neural mechanisms. Such inquiry has produced many key insights about memory, but there is widespread acknowledgement that little progress has been made in understanding precisely what is stored and how (Maguire 2022; Poeppel & Idsardi 2022).

The aim of this chapter is to explore whether there is a place for the memory trace in our theorizing about and investigation of memory. To do so, I set aside these acknowledged differences in how philosophers and memory scientists engage with memory traces, exploring instead how reflection on the concept could be generally beneficial. I propose stepping back to evaluate the various explanatory roles for which memory traces are invoked in both philosophical and empirical contexts. Making these background, often implicit lines of reasoning explicit provides an opportunity to subject them to scrutiny. We can consider whether these

explanatory roles are well motivated and well reasoned—and whether they in fact compel memory traces. We can also get a clearer sense of how memory traces are being understood, as distinct explanatory roles are likely to invoke distinct features. This provides a way of sorting amongst various philosophical views and their varying commitments. Whether a particular feature can be removed or diluted will depend, in part, on what role the trace is playing. It also makes evident the way that one's conception of the memory trace frames the empirical study of memory, shaping experimental design, investigation, and interpretation. Taking the time to make these guiding assumptions explicit and subject them to scrutiny can serve as an important reflection on experimental practice—perhaps suggesting reforms, but also identifying exciting new ways to move forward. To that end, the remaining discussion is guided by the following two questions:

1) What role are memory traces thought to play in an account of memory?
2) What must memory traces be like in order to play that role?

I divide exploration of these questions into *philosophy-first* and *science-first approaches*, covered in sections 2 and 3, respectively. In Section 2, I identify three standard lines of reasoning by which philosophers invoke memory traces, illustrating the distinct view of memory traces that comes from each. While none of the arguments is particularly successful in articulating a compelling and plausible account of memory traces, distinguishing amongst them may promote the development of further alternatives. The science-first approach in Section 3 highlights a way to identify such alternatives: from reflection on areas and eras of memory science where appeal to the trace concept is particularly useful. Here I focus on extracting an argument for memory traces from the recent resurgence of interest in the engram in neurobiology, arguing that it provides a novel and fruitful conception of the memory trace.

Philosophy-First Arguments for Memory Traces

Below I survey three general lines of reasoning that philosophers have used to argue for the existence of memory traces.[3] Each subsection concludes with a brief evaluation. As will be shown, none of the available accounts offer a particularly compelling argument for the existence of memory traces. Each faces significant challenges, both in terms of the coherence of the explanatory role proposed and in terms of the empirical plausibility of traces that have the features such a role requires. The final subsection, 2d, provides some reflections on this analysis and what it suggests as the way forward.

[3] The arguments identified in this section derive from my earlier survey of memory traces (Robins 2017). The current form involves some updates, most notably to ways of thinking about the features of memory traces corresponding to each explanatory role.

Representing the Past: Memory Traces as Mental Images

The first account of memory traces emerges from the need to explain our ability to represent past events or experiences. This route to traces can begin from a number of distinct philosophical commitments. One may, for example, subscribe to the view that all mental processes are representational, and so consider memory traces to be the particular form of representation required for memory. Augustine (2002) endorses a version of this view. This line of reasoning is best associated with the indirect realism of Locke (1975) and Hume (1978), according to which mental processes like perception and memory make us directly aware of representations and only indirectly aware of the features of the world that they represent. One could also have a more restricted sense of when mental representations are required and conclude that remembering requires memory traces, either because it has to be differentiated from perception (Martin 2001) or because remembering involves events or experiences that are no longer present or occurrent (Von Leyden 1961). Additionally, one could be led to this view of memory traces by considering the act of retrieval in particular. Retrieval involves selecting one past event from the set of such events retained in memory. The ability to do this requires explanation and so the memory trace is invoked to explain how retrieval is possible (e.g., James 1890). What must memory traces be like in order to play this role? While these arguments differ in scope, they converge on a view of memory traces as mental images: depictions of the prior events or experiences for which they stand in. Often, these are conceived of as visual, but could be images based in other sensory formats—a song's chorus, the musky scent of your grandmother's perfume. If the trace were not a replica of the past event, then it could not fulfill its role as the object of thought or goal of retrieval. I can, for example, remember my 8th birthday party. The event itself is not available for me, as an adult, to perceive. A representation of the event stands in. Similarly, when attempting to remember this birthday party, I may have the experience of scanning back through a host of mental images of childhood parties and events. My search stops, and remembering occurs, once I find the image that depicts my 8th birthday party. Importantly, on this view, memory traces are personal level entities.[4] That is, they focus on the role of traces in the act of remembering as it is experienced by the rememberer herself. The memory trace is what a person is aware of when thinking about the past. The memory trace of my 8th birthday party is the mental image that I 'see' in my mind, depicting the features of that experience I have retained.

The need to represent the past provides a role for memory traces that is easy to describe, but difficult to codify into an actual mechanism or process. In order for memory traces to play the role assigned to them, they must not only be mental

[4] The intended contrast here is with the subpersonal level. Philosophers draw the personal/subpersonal distinction in subtly distinct ways (e.g., Drayson 2014; Westfall 2024; Dennett 1969; Stich 1978). The sense appealed to here is broadly compatible with the range of views on offer, emphasizing the difference between what is available from a first-person, introspective perspective and what is only available from a third-person perspective.

images, but mental images that render their identity apparent. That is, a memory trace must not only depict the past experience, it must also make clear that what is being depicted (1) is of a particular experience (vs. some other experience) and (2) is a memory (vs. an imagination, hallucination, etc.). Each of these requirements poses a problem. Attempts to meet the first requirement make clear that the explanatory role of memory traces in retrieval is unstable. The argument is meant to show that retrieval requires a memory trace. But how is it that I know which trace to retrieve? That is: even if it 'looks like' my 8th birthday, how do I know what my 8th birthday looks like such that I can recognize it amongst images of other past events? If I know what it looks like already, then the initial motivation for the memory trace is undercut. A memory trace is not required for retrieval, as the argument initially supposed; the work is being done instead by knowledge I already possess. The other alternative thus seems preferable: it is not that I know in advance what I want to remember, but that I remember how to navigate to that item in my memory or recognize it upon presentation. But this route also leads to problems. The argument began from the claim that remembering requires a trace. In explaining how the trace is identified, appeal is made to another instance of remembering—remembering where this trace can be found or what features to look for. This remembering, too, must require a trace. A memory of how to locate or recognize a trace would be required for each trace. So the question can be raised again for this trace: how is the correct one identified? And so on. The need to account for each subsequent appeal to remembering presents the dilemma anew. Either this continues indefinitely, an infinite regress through which the traces required for a single instance of remembering pile up, or it stops with a form of remembering that does not require the retrieval of a trace—an abandonment of the central claim from which this began.[5]

Even if this line of argument were more successful, there is independent reason to be skeptical of the view of memory traces it entails. Consideration of what representing the past entails leads to a view of memory traces as mental images that are recognizable as depictions of past experiences. But there is significant reason to doubt the existence of any such identifying marker. Specific proposals are often subject to targeted criticism, as in Reid's (2002) challenge to Hume's (1978) claim that 'vivacity' is what distinguishes memory images from those of perception, imagination, and the like. The more general concern derives from decades of empirical research. It is not simply that the feeling of remembering is hard to render precise, but that the various candidate markers are licentious—often accompanying inaccurate depictions of the past or representations of events that never occurred (e.g., Bernstein & Loftus 2009).

[5] A version of this argument can be found in Heil (1978).

Sustaining Causation: Memory Traces as Filling the Gap

The second account of memory traces begins with reflection on a widely accepted claim: Remembering is a diachronic process. It involves an initial event or experience that is recalled during a subsequent experience. Memory traces are invoked as the intermediaries required to explain this temporally extended process. By the middle of the twentieth century, it was especially common to understand remembering's diachronic nature in causal terms—i.e., as seeing the subsequent recall of an event like one's 8th birthday party as an effect of the original event, the 8th birthday party itself.[6] From here, the commitment to memory traces falls out of an understanding of the nature of causation and what is required for sustaining a diachronic causal process.

There are many theories of causation available—and correspondingly, many particular ways to generate the memory trace commitment from one's view of causation. The commitment could arise from a general constraint on the nature of causation, specifically from the denial of action at a distance. Suspicion of remote causes is frequently found in discussions of causation, as in Leibniz's (2000) critique of Newton's invocation of gravity as an appeal to occult forces. Remembering is, in this respect, much like gravity: alleged to provide a causal influence on events from which it is spatiotemporally distant. Appeal to the memory trace as an intermediary precludes any accusations of the occult. On the broadly Humean view of causation, causes must precede their effects and immediately so. Otherwise, there is a temporal gap by which the presumed or intended cause could be pre-empted. Memories may be effects of experience, but they rarely if ever follow directly from an experience. An intervening trace, formed as the result of experience and sustained until the time of remembering, provides the requisite contiguity (Bernecker 2008). The causal commitment to memory traces could also emerge from within a process view of causation (e.g., Salmon 1984; Dowe 1992), where causal relations are understood in terms of conserved quantities or mark transmission. On such a view, the memory trace would be the quantity conserved or the mark transmitted. Across these distinct views of causation, a core commitment emerges—the need for a memory trace to sustain the temporally extended process of remembering.

To play this role, a memory trace must be the kind of thing that can participate in causal relations—i.e., it must be physical.[7] It must also have features that support its intermediary, connective task. Cases of remembering can occur long after the experiences from which they derive, demanding the trace cover an extended

[6] Commitment to this causal understanding of memory can be seen, for example, in Ayer (1956), Martin and Deutscher (1966), Shoemaker (1970), Anscombe (1981), and Armstrong (1987).

[7] Within this generic physicalist commitment, particular views of causation characterize causal relata differently, which could result in a number of distinct particular commitments: to the memory trace as an object, event, fact, conserved quantity, etc.

temporal gap. The trace must therefore be the kind of physical thing that can persist across time. It must be stable, perhaps even static—capable of preserving the past experience's causal influence over extended periods, without disruption or deterrence from other influences.

The success of this argument for memory traces is dependent upon the success of these ways of thinking about the nature of causation. This is already suspicious. It would be preferable for the commitment to memory traces to derive from consideration of what remembering requires. Tying its fate to a particular way of thinking about causation opens up the possibility that the need for memory traces will dissolve if/when an alternative way of thinking about causation emerges. Indeed, this has happened. Concerns about action at a distance are less compelling given the kinds of things that are now thought to be capable of standing in causal relations. Moreover, there are counterfactual (Lewis 1973) and manipulability (Woodward 2003) theories of causation that lack contiguity, conservation, or transmission constraints. Such approaches to causation are increasingly popular, especially amongst those working in the biological and social sciences where processes like remembering are investigated.

There are additional problems with the memory traces themselves, as they are understood on this account. The memory trace earns its place by being stable and static, providing a past experience with a direct route toward expression in the present. But the brain in which such traces are presumed to reside seems an unlikely host. Neuroscientists are increasingly interested in thinking about the brain as a fluid system, better understood through dynamics and probabilities than stable structures and serial processes (e.g., Gerstner, Kistler, Naud, & Paninski 2014). As Lynn Nadel (2007) has noted, this field-wide focus entails the 'demise for the fixed trace' (p. 181). There is, more generally, a lack of clarity in understanding how the contiguity requirement on remembering should be applied to the neural processes that underlie it. How close is close enough? One could, for instance, worry about the insistence on spatiotemporal contiguity and the endorsement of the nearly consensus view of memory traces as synaptic. Synapses are, after all, *gaps*. It's hard to take this seriously as an objection; it seems both unfair and uninformed about neural systems. But even so it does make clear that which kinds of gaps in spatial and temporal relations are regarded as significant is likely to vary with both the systems under investigation and our interests in studying such systems. Contiguity itself does not provide much of a guide for identifying or understanding memory traces, or any other causes.

Remembering vs. Relearning: Memory Traces as Internal States

A final role for traces emerges from a tried-and-true philosophical method of conceptual analysis. Much of contemporary philosophy in the English-speaking world involves the analysis of linguistic predicates (i.e. concepts) and how to define them,

or capture all and only their features or requirements.[8] Here the claim is that an analysis of *remembering* reveals memory traces to be a necessary feature, component, or requirement. Memory traces are simply built in to what it means to remember. The analysis proceeds via the *method of cases* (Machery 2017). One considers a range of scenarios, often hypothetical, where candidate features of the concept under scrutiny are lacking or altered. Responses to the case—e.g., determination of whether the scenario presented involves remembering or not—are used to establish what the concept requires.[9]

Martin and Deutscher's (1966) *Remembering* is a primary example of this method, applied—as the title of their paper suggests—to remembering. Subsequent work by Bernecker (2010) and Debus (2010) add further cases to the analysis. Martin and Deutscher begin their analysis from the presumption that remembering requires, at a minimum, an accurate representation of something that has happened to the would-be rememberer. In order to remember my 8th birthday party, it must first be the case that I had such a party and also that my thinking about it now captures what happened.[10] They then go on to generate a range of hypothetical cases where both of these conditions are met and yet remembering does not occur. The key case that they consider is one of relearning: a case where a person learns something, forgets it, and then reacquires it from another source. Their example involves a person named Kent, who is in an accident and later suffers amnesia that causes him to forget the accident (and other events). Kent is then provided with information about the event from another source, allowing him to re-represent it. In such a case, it is proposed, Kent has *relearned*, not *remembered,* the event.[11] To account for the difference between remembering and relearning, remembering must have a feature that is missing or malformed in the relearning case. The feature that is missing in cases of relearning is the memory trace. In cases of relearning, the memory trace has been lost. In cases of remembering, it has been retained.

So what would a memory trace have to be like in order to support this distinction? The case of relearning is designed so that it matches remembering on a number of features. Kent's case is meant to be like remembering in the existence of the remembered event (the accident), the accuracy of the current thoughts about that event—and even in the causal connection between the past event and the current thoughts about it.[12] The key difference between the two cases is the relation of that causal chain to the would-be rememberer. In remembering, but not

[8] This chapter, with its focus on what memory traces could be, is hardly an exception.

[9] Whether responses to thought experiments should be understood as intuitions, reasoned judgements, etc. is a matter of current debate in philosophy. Nothing in the brief characterization here should be taken as favoring any particular view on this issue.

[10] Martin and Deutscher do not provide much further elaboration on what accuracy requires. There are a host of further questions one can press about this requirement (e.g., are omissions and commissions equally problematic? How much must be represented in order to achieve accuracy?), but I am setting them aside for present purposes.

[11] When psychologists talk about relearning, they are often referring to a slightly different phenomenon: the re-exposure to information when one has not previously forgotten that information. The 'savings during relearning' paradigm can be traced to Ebbinghaus (see Nelson 1985).

[12] In this way, Martin and Deutscher's analysis of remembering includes commitment to a causal condition on remembering. It could be possible for someone to develop an alternative analysis that advocates for distinguishing remembering from relearning without the causal condition.

in relearning, the causal chain stays *inside* the person, in some intuitive, but difficult to articulate sense. The distinction is best understood, I have argued (Robins 2016), as a *cognitive* difference. It is an assessment of the causal history of the current thought about the past event. Is the capacity to produce it something that has been retained or something that has been regained? To be remembering, that capacity must reside inside the cognitive system that produces the thought being assessed.

This of course gives rise to questions about the mind's perimeter. The cognitive boundary has been challenged from many directions in recent decades, as philosophers and cognitive scientists have argued that the mind extends out into the world and other people, as well as down into various parts of the body. Luckily, differentiating remembering from relearning does not require settling this issue. Wherever that boundary is drawn, remembering will fall inside it and relearning outside. Memory traces are internal states that produce instances of remembering.

The biggest challenge to this argument for memory traces is the number of philosophers who are not convinced by consideration of relearning. The case of Kent, for some, fails to elicit the presumed response. Kent is remembering—or at least the possibility that Kent is remembering is not ruled out by the conditions as stated by Martin and Deutscher (Michaelian 2016). It is further unclear whether the distinction between remembering and relearning matters beyond contrived philosophical contexts. There are some instances where it seems to matter: the distinction between testimony and hearsay, or a perfect exam that results from studying versus a cheat sheet. But in many cases, even if the difference is understood, it is impossible to tell. As Martin and Deutscher (1966) themselves note, many recollections from early childhood are like this. We often do not know whether events from our early years are ones we remember or ones we remember the retelling of from others.

Even if this line of argument in favour of memory traces were particularly compelling, it does not provide an especially rich account of memory traces. It tells us only that they are states internal to the cognitive system, derived from past experiences.

So What Now?

Above I identified three distinct lines of argument that lead to a commitment to memory traces, each yielding a distinct conception of what traces are like: the memory trace as mental image, the memory trace as stable causal force, and the memory trace as internally maintained state. None of the features identified were in and of themselves surprising: representational, causal, and historical features are familiar to accounts of memory traces, as the definitions from the introduction illustrate. What is interesting, however, is how the features align with particular arguments.

While these features are often lumped together in discussions both for and against traces, the above analysis makes clear how distinct features are connected to distinct ways of arguing for traces.

For example, the representational, imagistic features of memory traces are essential for accounts concerned with representing the past. Such views need not have any additional requirement that memory traces be internal or supported by the right kind of causal linkages. Representational views could even deny that remembering is causal. Conversely, views that derive from a commitment to remembering as causal could eschew any commitment to representational or imagistic features and characterize traces as merely sustaining a causal connection (e.g., Werning 2020). Even the view of memory traces that derives from considering cases of relearning can be formulated without the features that stem from the other two lines of argument. The internal stages need not support mental imagery. And while it is true that Martin and Deutscher's (1966) analysis does presuppose that remembering is causal, this need not lead to a requirement that that traces be static and stable. They could simply endorse one of the available views of causation that does not preclude spatiotemporal gaps.

The above evaluation demonstrates the importance of excavating the line of reasoning being used to defend the idea of memory traces, as it is critical for determining which of the myriad features often associated with memory traces are needed for the argument at hand. Unfortunately, it also demonstrates something else: none of these lines of argument is particularly successful. Each is ineffective at securing a clear and compelling explanatory role for memory traces. And, in each case, there are also independent reasons to be suspicious about the existence of memory traces with the requisite features. In searching for a place for the memory trace, it seems that it would be more profitable to go in a new direction.

Science-First Arguments for Memory Traces

In this section, I turn to what I loosely describe as science-first ways of arguing for memory traces. As stated in the introduction, the memory trace often serves as a background commitment for empirical investigation of memory mechanisms. There are times, however, when the concept receives more explicit attention. Although the focus in such times is not generally on arguing for memory traces, such an argument may be identifiable from the way in which the interest in and usefulness of the trace emerges.

To this end, I am particularly interested in the resurgence of work on the engram in contemporary neurobiology. Elsewhere I have argued that this line of research marks a significant change in the kinds of inquiry into memory mechanisms that neurobiologists of memory are conducting (Robins 2018; Robins 2023). Here I focus on the motivation for that change and how it might be built up into an argument for memory traces. In 3a, I discuss engram research and its significance in the current

context. In 3b, I identify the explanandum and how it might be developed into an argument for memory traces. In 3c, I assess its explanatory adequacy in comparison to other approaches.

The Engram Renaissance

The neurobiology of memory is currently experiencing an 'engram renaissance' (Josselyn, Köhler, & Frankland 2017: 4647). Researchers working in this area do not regularly refer to their work in this way, and indeed the use of this phrase in the paper just cited is something of an offhand remark. I have amplified it here because it is apt for characterizing the significant change the field is undergoing. As the term *renaissance* implies, the current fervour does not involve a new idea or concept, but the rebirth of an old one. 'Engram' was coined by the zoologist Richard Semon (1904/1921) more than a century ago. The term has been in use since, but only intermittently. A *Web of Science* search shows that Semon's book was cited more in the decade from 2010 to 2020 than in the previous ten and half decades since its initial publication.[13] *Renaissance* also implies a time of new discovery and the exploration of themes and connections that disrupt established boundaries of inquiry. In this way, research into the engram has been exciting not only for what it reveals about the basic mechanisms of memory, but for the opportunities it provides to connect with broader areas of memory science: investigations of false memory in cognitive psychology, models of long-term memory consolidation in neuropsychology, and the treatment of Alzheimer's.

To make clear how the engram renaissance is relevant to this investigation of memory traces requires saying more about both the engram and recent discoveries. First, the engram. Semon (1921) defined the engram as the physical or chemical change to the brain that results from learning. Semon coined the term to guide inquiry, not to label a discovery. In this way, the engram is a more technical, scientific version of the memory trace. How much overlap there is between the two terms depends on how widely one's scope of inquiry is drawn. As we saw above in Section 2, the memory trace can be understood in many ways, not all of which invoke physical features, much less the brain. For philosophers, then, the two concepts can diverge significantly. Amongst memory scientists who are focused on cognitive and neural systems, however, the two are generally treated as synonyms. Insofar as the engram is a way of conceiving of memory traces, here I will follow the scientific practice of treating 'memory trace' and 'engram' as interchangeable.

In the neurobiology of memory, the engram has long persisted as something of a regulatory ideal: a commitment that guides inquiry but is not taken too literally. That is, the investigation of the mechanisms of memory is framed around the presumption that memory involves the encoding, storage, and retrieval of engrams,

[13] Eighty-three vs. twenty-nine. Search conducted January 2021.

but studies are designed around the general investigation of these processes, not the search for individual engrams. With the development of new experimental tools, this has changed. Advances in activity-dependent cell-labelling (Mayford 2014), in vivo calcium imaging (Yang & Yuste 2017), and chemogenetics (Armbruster et al. 2007) have brought increased precision to the investigation and manipulation of memory mechanisms. But the real driver of change has been *optogenetics*, a technique for controlling neural activity with light (Deisseroth 2010). Optogenetics allows neuroscientists to target a selective group of neurons and activate or inhibit their expression in living, behaving organisms.

In application to memory, this has meant the ability to create light-responsive engrams (Liu et al. 2012). Working mostly with mice, researchers have shown that optogenetics can be coupled with standard spatial memory paradigms so that the neurons active during the encoding of a spatial memory will subsequently express a light-sensitive protein. The expression of this protein makes it possible to reactivate the neurons that encoded the memory by exposing them to light. Memory retrieval via light switch.[14] The availability of this new form of intervention affords researchers a host of opportunities to explore the features of memory formation, storage, and reactivation under a host of conditions. This has led to a range of fascinating discoveries, including: the production of false memory in non-human animals (Ramirez et al. 2013); the ability to change the valence of a memory from positive to negative and vice versa (Redondo et al. 2014); recovery of memory in early stage Alzheimer's models (Roy et al. 2016); and the creation and implantation of an artificial engram (Vetere et al. 2019).

My interest here is not in a direct assessment of this research programme, nor the validity or significance of any of these results. Regardless of how one views this line of research, its growth and prominence is undeniable. My interest is in the role of the engram/memory trace within this research programme. Why is this new approach to investigating memory re-exciting talk of the engram? It cannot be written off as simply the result of a new investigative tool or discovery. After all, there have been many such tools over time. Viewed from a wider lens, the broad consensus that now exists about the mechanisms of memory is the result of a steady stream of tool development and discovery across several decades and areas of neuroscience (Craver 2003; Silva, Landreth, & Bickle 2014). My question is: why is the engram concept of such interest, *now* as opposed to other significant moments in the history of memory investigation? Why did it not occur alongside the development of fMRI or CRISPR?

I want to suggest that the answer to this question has to do with the specific kind of change that optogenetic intervention has brought about. There has been a shift, from investigating *mechanisms* to investigating *memories*. The neuroscience of memory has long been a study of memory's mechanisms (Craver 2007). How are long-term

[14] A more thorough review of the method can be found in Robins (2023).

memories formed? What is the nature of consolidation (and reconsolidation)? These questions are answered by identifying neural mechanisms, organized clumps of entities and activities, and by characterizing how their interactions change under a range of conditions, perturbations, and interventions. Now, with the advent of optogenetics, researchers can use this general mechanistic understanding to explore the formation, persistence, and modification of particular memories. That is, optogenetics allows for the identification, investigation, and manipulation of token activities in the memory system.

This provides an answer to the question with which this section began: the memory trace/engram becomes relevant and interesting when researchers have the capacity to identify, isolate, and manipulate memories of particular past experiences. As the available tools make it possible to study the retention of specific, token occurrences, the idea of the engram returns. Can this be used to develop a new line of argument for memory traces—an argument built around their role in explaining the retention of particular past experiences? I explore this in the next subsection below.

Retaining Particulars

First, we must determine whether the retention of particulars is a phenomenon worth explaining. If it rarely occurred, or occurred only in limited or contrived contexts, or was marginal in some other way, then this would not be a particularly promising line of inquiry.

We do in fact remember particular past encounters—at least sometimes. I will substantiate this claim with evidence below, but its truth should (I hope) seem fairly clear to most readers simply from experience. It occurs in cases of one-shot learning and instances of episodic remembering. Of course, not all learning happens this way and not all attempts at episodic remembering are successful. What matters for present purposes is that it does happen. The scope of 'we' in this claim is intended to be vague and somewhat broad. I have in mind not only humans, but also other animals whose memories are well studied (e.g., mice, scrub jays), possibly more. I do, however, take 'remember' to be something of a success term here. The intention is to highlight cases where we get it right, without worrying (at least for now) about how to set specific standards of accuracy and correctness. 'Particular past encounters' is a mouthful, but the aim of such a term is to remain neutral amongst memory taxonomies and ways of defining episodic or event memory while still highlighting token occurrences. How tokening occurs, how the boundaries of tokens are set, etc. are interesting puzzles, but not ones that need to be settled for establishing this basic claim about retaining particulars. The spatiotemporal boundaries of token occurrences are likely to be set by mechanisms of event segmentation, which may vary across species, individuals, and/or contexts.

The retention of particular past experiences is something that a theory of remembering ought to be able to explain. Retaining specific facts, ideas, and experiences is a quintessential form of memory. When asked to think of a memory, such occurrences are often what's selected. It is an emblematic form of memory, arguably where it's most distinct from other capacities like perception and belief. Appeal to a memory, or engram, provides a straightforward explanation of it: features of the occurrence are encoded, retained, and later reactivated. Whether such an explanation is compelling enough to support an argument in favour of memory traces depends on whether there are other ways of explaining the retention of particulars and, if so, their comparative strength. That is, the appeal to memory traces to explain the retention of token occurrences could amount to a successful argument for the existence of memory traces if it can be shown to be the best possible explanation of this phenomenon.[15]

Successful retention of particulars has not received a lot of direct attention in the past several decades. Instead, explanatory frameworks have been centred around (1) memory errors and (2) the role of patterns, generalizations, and abstractions in remembering processes, both as a way of accounting for (1) and in attempt to align memory with theories of cognitive architecture more generally.

Approaches to memory mechanisms that highlight their role in broader cognitive and neural processes, like those identified in the introduction, are prime examples. Such views are at pains to explain how memory errors can occur as frequently and systematically as they do. Dan Schachter, a leading proponent of the episodic simulation hypothesis, describes how evidence of false memory has shaped his thinking about the nature and function of memory more generally, saying that 'it makes little sense that evolution would have yielded such a deeply flawed system' (2019: 265). Instead, he and others have come to support an alternative way of looking at memory systems, such that memory errors are viewed 'as byproducts of otherwise adaptive features of memory' (Schacter 2019: 265). Similarly, philosophers have recommended reorienting our thinking about the relation between memory's failures and successes such that they can both be understood as cases of the memory system 'doing what it is supposed to do' (De Brigard 2013: 172).

What is it that memory is supposed to be doing? What are its adaptive features? There are a range of distinct positive proposals on offer. They all supplant the standard approach to memory as a faculty for retention, and with it the encoding-storage-retrieval approach through which memory traces may play a role. They highlight instead a broader cognitive aim—e.g., prediction, inference, decision-making, narrative, or schematic cohesion—and posit a more generalized and distributed form of information storage suited to this aim. Addis (2020), for example, thinks of constructive episodic simulation as a critical cognitive capacity.

[15] De Brigard (2020) offers a similar Inference to the Best Explanation argument for memory traces, where what's being explained is the causal connection between an experience and its subsequent recollection.

Constructive episodic simulation is here understood as the ability to generate representations of what the world is like from a first-person (episodic) perspective across temporal and conceptual dimensions. I can construct episodic simulations about what I will do tomorrow, or next month, or ten years from now, or in retirement. I can also construct episodic simulations of what I should have done yesterday, or what I would have done if circumstances had been otherwise, or what I would do now with difference resources, or what I could do if anything were possible. To support these simulations, the underlying cognitive system is organized to support schematic representations and associations between concepts and other event components. Schemas provide general, thematic structures for events and activities and associationist networks provide further information about things and features that tend to co-occur. Together, these can be used to build compelling episodic simulations of events—ones that should happen, might have happened, or could happen in the future.

Remembering what did happen in the past is, on this view, yet another particular application of this general capacity. Event schemas and associated features are used to build a plausible representation of the past. This works reasonably well. When the events we are trying to remember fit with these schemas and associations, the resultant simulations get it right, or mostly so. It also explains how memory errors occur, as the influence of schemas and associations may lead to the insertion of themes and elements that were not a part of the past event in question. While the influence on remembering is bothersome, it is useful for other forms of episodic simulation, and so understandable as a feature of this system.

Addis's (2018, 2020) view is only one example, but it is reflective of a general explanatory strategy that de-emphasizes traces. No one denies that successful remembering occurs. And gestures are often made to suggest how such cases could be explained in terms of these overall frameworks. Still, the focus is elsewhere—on broader cognitive tasks and the architectural principles by which cognitive systems achieve them. By highlighting the retention of particulars, work on the engram in neurobiology offers an opportunity to direct some of our attention back to this phenomenon and whether it is explained well enough on such approaches. I take this up in the next section.

Memory Traces and Explanatory Adequacy

Some views of memory are focused on accounting for memory errors and/or memory as a sub-component of more general cognitive processes. On such approaches, successful retention of particulars is not denied, it is just not given much attention. The lack of attention has consequences, which challenges the explanatory adequacy of such proposals.

This is well illustrated in a recent study by Diamond and colleagues (2020) that explores the accuracy of event memory in real-world contexts. They measured memory for two 'complex and immersive yet verifiable real-world experiences' (p. 1545): (1) a training on mask-fitting procedure given to hospital employees, and (2) a tour of a hospital foyer with lots of distinctive art and architecture. Importantly, participants were recruited to the study *after* their respective experience, removing concerns about experimental context and participant expectations from the encoding process. Once participants were recruited to the study, the researchers used free recall techniques to elicit memories and measure how much of the initial information had been retained, at intervals ranging from two days (for the hospital tour) to two years (for the mask-fitting training). Across both experiences, they found that participants were able to provide many event-specific details in their recollections—and, importantly, the details provided were highly accurate. For example, participants who experienced the mask-fitting training were able to generate an average of forty to fifty details about the experience, more than 95% of which were confirmed as accurate, even when tested one to two years after it occurred.

Alongside these measures, Diamond and colleagues (2020) conducted a survey of memory scientists and other academics, where they presented survey participants with hypothetical cases modelled on the real-world experiences assessed above, asking them to estimate how accurate a person's recall would be in such a scenario. Survey responses showed low confidence in people's capacity for accurate retention. When asked, for instance, to estimate how well a healthy 30-year-old would perform when recalling the details of a museum tour two days after the experience, the researchers surveyed estimated that accuracy would be between 20 and 30%. For the hospital tour participants in this age range, accuracy was around 94%.

The Diamond et al. (2020) study demonstrates that successful retention of particular past experiences occurs and illustrates an ingenious way of uncovering it. The study simultaneously shows the problems that can arise when theoretical and empirical inquiry is too focused on memory errors. As they state, the results show that many researcher's 'views of human memory accuracy are overly pessimistic' (2020: 1552). This finding matters not only because of the disconnect it reveals between everyday remembering phenomena and researchers' suppositions about it, but also because these expectations are ones that researchers are unlikely to correct if they are not motivated to investigate. One's estimate of how likely successful retention is will shape one's expectation of whether it is fruitful to explore empirically or incorporate into one's theoretical framework.

With these expectations of memory's accuracy challenged, we are then free to explore and notice other domains where retention of particulars occurs. This could include cases of highly superior autobiographical memory (LePort et al. 2017) or broader investigations of expertise, both memory-specific (Foer 2011) and more generally (Ericsson & Lehman 1996). We might also find cases in less obvious places,

like models of PTSD (Zhang et al. 2020) and addiction (Chiew & Adcock 2019), where the retention of particulars has less positive consequences.

This may also, in turn, lead to critical reflection on the explanatory adequacy of false memory frameworks for the very phenomena they were developed to explain. Eyewitness testimony has, for example, been one of the key forms of false memory to generate attention and concern. Evidence of the ease and extent to which first-hand reports of past experiences can be manipulated and distorted has challenged how many think about the significance of testimony, challenging its long-standing pride of place in legal contexts (Loftus 2003 for review). While the effects of interrogation techniques on testimony are clear and well documented, their overall prevalence and significance may have been overstated. As Wixted and colleagues (2018) argue, our understanding of how errors can come about in testimony is consistent with a view that such testimony can be accurate when such forces are kept at bay.

How troubled one is by cases that do not fit the general pattern of explanation critically depends on how many cases one thinks there are. If there are only a handful of such cases, then even the realization that they are not captured well, or at all, by one's framework may not be particularly concerning and could possibly be absorbed with little to no alteration to one's theoretical commitments. A view of remembering as supported by a cognitive system of schemas and associations, like Addis's (2020) view above, could likely take on board a few non-schematic, individual event representations. As more such cases need to be accommodated, this becomes more difficult and continuing to absorb the anomalies becomes inadvisable. Taking on too many exceptions risks negating the explanatory virtues of the proposed architecture. Appeal to schemas is meant to provide an account of how information can be efficiently and effectively organized. As the amount of non-schematically organized information grows, the ability for schemas to play their organizational, streamlining role is complicated. So too for systems built around principles of association, frequency, narrative cohesion, etc. The power and explanatory benefit of such organizational frameworks comes from their breadth. This is not, of course, to say that schemas have no role to play in scaffolding particular memories. It seems plausible that they play a key role in preserving content, facilitating pattern completion, etc.[16]

This criticism should not be understood as a recommendation that such cognitive frameworks be rejected. Schematic, associationist, and frequentist processes clearly play a role in many cognitive operations. The question, as I see it, is how to understand the scope of their explanatory power. Should they be used to explain all processes and operations? If we identify phenomena/systems/processes that do not fit well into these frameworks, should we attempt to subsume them or make space for distinct, complementary architectures? My preference is to support the latter. Setting constraints on the applicability of these pattern-based processes would be a

[16] Thanks to Sara Aronowitz for pressing this point.

way to retain their explanatory virtues, while allowing for alternative explanation of phenomena that are ill-suited to that framework. What the details of such a proposal (likely multiple proposals) would look like is still far from clear. But it's still progress to recognize the work that needs to be done.

To summarize this section, I have used the case of the engram renaissance in neurobiology to argue that the memory trace is a useful concept for memory scientists when they are exploring the retention of particular past experiences. In reflecting on this phenomenon of retaining particulars, we can come to see that it is an important, central phenomenon to be explained. As the situation stands currently, trace-based explanations of our ability to retain particular past occurrences fare better than alternatives focused on the role of more general cognitive processes. It is possible that this apparent advantage is illusory. As noted above, the need to explain retention of particulars does not receive much direct attention. It could be that, once they do so, more plausible and competitive explanations will emerge. For now, there is at least a place for the memory trace worth investigating further.

Memory Traces as Discrete Entities

The last section sketched an explanatory role for memory traces: as the best available explanation of our ability to retain particular past occurrences. We can now go on to ask the second question: what must memory traces be like in order to play this role? Here I elaborate, briefly, on the features such a view of memory traces compels.

To account for the ability to retain particular past occurrences, memory traces must be discrete entities. Discrete here stands in opposition to continuous. The form of retention memory traces support must be one that allows for particular past encounters to be retained in ways that can, at least sometimes, survive interference from retention of other particular past encounters and more general forms of information storage, processing, and updating. In other words, if retaining information from a particular experience is explained by a trace, then the trace that makes that possible has to be built such that it is capable of remaining distinct from other traces and from other non-trace kinds of information.

This commitment to discreteness places a constraint on both the structure and content of the memory trace. Discreteness is, in fact, a structural commitment in the sense that Hebb (1949) first proposed as a constraint on the engram, as an intended contrast with diffuse and dynamic patterns of activity. The claim that memory traces are *discrete* should not be mistaken for the claim that memory traces are *local*. Memory traces could be distributed, in a sense operative at multiple levels of organization: across networks, regions of the brain, neural populations. Similarly, traces can be *discrete* without being *static*. The kind of retention they support must be capable of resilience amid other instances and form of retention, but the precise vehicles by which that resilience is supported could change over time. In order for discrete traces to explain the retention of particular past encounters, they must

not only contribute causally to their remembering but do so in a way that is reflective of the content acquired or encoded from that encounter. They must, in other words, be information-bearing or representation-supporting in some critical way. How this content is to be understood is not fully clear. There is likely room for a range of proposals as to what this content is and how it is carried by the trace.

Thus, even with this commitment to the memory trace as a discrete entity, there are a range of ways in which the constraints on its structure and content could be understood. And from here, even more work to be done articulating how their structure and content are related to one another—both at any particular moment, and over time. Exploring various proposals would offer an increasingly rich understanding of what memory traces are or could be, perhaps supporting multiple distinct forms or competing proposals, the testing of which could yield fruitful new lines of theoretical and empirical inquiry.

Conclusion

The memory trace has a long and complicated history. Whether interest in the memory trace should be restricted to retrospective storytelling is, for the moment, an open question. Many would happily leave it behind as memory research goes forward. The recent development of optogenetic tools has led at least some memory researchers to consider the concept worth dusting off and carrying forward. The role they see for memory traces, in the retention of particular past encounters, is a promising one—and one that is otherwise neglected in current theorizing. How far such a proposal can go remains to be seen. Regardless of the outcome, exploring whether there is a place for the memory trace provides a critical moment of reflection on our ways of thinking about the nature and purpose of remembering.

References

Addis, D.R. (2018). Are episodic memories special? On the sameness of remembered and imagined event simulation. *Journal of the Royal Society of New Zealand*, 48, 64–88.

Addis, D.R. (2020). Mental time travel? A neurocognitive model of event simulation. *Review of Philosophy and Psychology*, 11(2), 233–59.

Anscombe, G.E.M. (1981). Memory, experience, and causation. In G.E.M. Anscombe (ed.), *Collected Philosophical Papers, Vol. II: Metaphysics and the Philosophy of Mind* (pp. 120–30). Oxford, UK: Oxford University Press.

Aristotle. *De Memoria et Reminiscentia* (W. D. Ross, trans, 1955). Oxford, UK: Clarendon Press.

Armbruster, B.N., Li, X., Pausch, M.H., Herlitze, S., & Roth, B.L. (2007). Evolving the lock to fit the key to create a family of G protein-coupled receptors potently activated by an inert ligand. *Proceedings of the National Academy of Sciences*, *104*, 5163–68.

Armstrong, D. M. (1987). Mental concepts: the causal analysis. In R.L. Gregory (ed.), *The Oxford Companion to the Mind* (pp. 464–65). Oxford: Oxford University Press.

Augustine of Hippo (2002). *On the Trinity: Books 8–15*. Trans. S. McKenna. Ed. G.B. Matthews. Cambridge: Cambridge University Press.

Ayer, A.J. (1956). *The Problem of Knowledge*. Harmondsworth, UK: Penguin.

Bernecker, S. (2008). *The Metaphysics of Memory*. Dordrecht, Netherlands: Springer.

Bernecker, S. (2010). *Memory: A Philosophical Study*. New York: Oxford University Press.

Bernstein, D.M., & Loftus, E.F. (2009). How to tell if a particular memory is true or false. *Perspectives on Psychological Science*, 4, 370–74.

Buszáki, G. (2019). *The Brain from Inside Out*. New York: Oxford University Press.

Chiew, K., & Adcock, A. (2019). Motivated memory. In K.A. Renninger & S.E. Hidi (eds.), *The Cambridge Handbook of Motivation and Learning* (pp. 517–46). Cambridge: Cambridge University Press.

Craver, C. F. (2003). Interlevel experiments and multilevel mechanisms in the neuroscience of memory. *Philosophy of Science*, 69, S83–97.

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

De Brigard, F. (2013). Is memory for remembering? Recollection as a form of episodic hypothetical thinking. *Synthese*, 191(2), 155–85.

De Brigard, F. (2020). The explanatory indispensability of memory traces. *The Harvard Review of Philosophy*, 27, 23–47.

De Brigard (this volume). Simulationism and Memory Traces. In S. Aronowitz & L. Nadel (eds.), *Space, Time, and Memory*. Oxford University Press.

Debus, D. (2010). Accounting for epistemic relevance: A new problem for the causal theory of memory. *American Philosophical Quarterly*, 47, 17–29.

Deisseroth, K. (2010). Optogenetics: Controlling the brain with light. *Scientific American* 303(5), 48–55.

Dennett, D.C. (1969). *Content and Consciousness*. Cambridge, MA: MIT Press.

Diamond, N.B., Armson, M.J., & Levine, B. (2020). The truth is out there: Accuracy of recall of verifiable real-world events. *Psychological Science*, 31, 1544–56.

Dowe, P. (1992). Process causality and asymmetry. *Erkenntnis*, 37, 179–96.

Draaisma, D. (2000). *Metaphors of Memory: A History of Ideas About the Mind*. Cambridge: Cambridge University Press.

Drayzon, Z. (2014). The personal/subpersonal distinction. *Philosophy Compass*, 9, 338–46.

Dudai, Y. (2007). Memory: it's all about representations. In H.L. Roediger, Y. Dudai, and S.M. Fitzpatrick (eds.), *Science of Memory Concepts* (pp. 13–16). Oxford: Oxford University Press.

Ericsson, K.A., & Lehman, A.C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology*, 47, 273–305.

Foer, J. (2011). *Moonwalking with Einstein*. New York: Penguin Press.

Gerstner, W., Kistler, W. M., Naud, R., & Paninski, L. (2014). *Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition*. Cambridge: Cambridge University Press.

Hebb, D. (1949). *The Organization of Behavior*. New York: Wiley & Sons.

Heil, J. (1978). Traces of things past. *Philosophy of Science*, 45, 60–72.

Hume, D. (1978). *A Treatise of Human Nature*, ed. L.A. Selby-Bigge, 2nd ed. Oxford: Clarendon Press.

Hutto, D., & Peeters, A. (2018). The Roots of Remembering. In K. Michaelian, D. Debus, & D. Perrin (eds.), *New Directions in the Philosophy of Memory*. Routledge.

James, W. (1890). *The Principles of Psychology*. London: Macmillan.

Josselyn, S.A., Köhler, S., & Frankland, P.W. (2017). Heroes of the engram. *Journal of Neuroscience*, 37, 4647–57.

Leibniz, G.W. (2000). *G. W. Leibniz and Samuel Clarke*: *Correspondence*. Hackett.

Leport, A.K.R., Stark, S.M., McGaugh, J.L., & Stark, C.E.L. (2017). A cognitive assessment of Highly Superior Autobiographical Memory. *Memory*, 25, 276–88.

Lewis, D. (1973). Causation. *Journal of Philosophy*, 70, 556–67.

Liu X., Ramirez, S., Pang, P., Puryear, C., Govindarajan, A., Deisseroth, K., & Tonegawa S. (2012). Optogenetic stimulation of a hippocampal engram activates fear memory recall. *Nature*, 484, 381–85.

Locke, J. (1975). *An Essay Concerning Human Understanding*, ed. P.H. Nidditch. Oxford: Oxford University Press.

Loftus, E. F. (2003). Our changeable memories: Legal and practical implications. *Nature Reviews: Neuroscience*, 4(3), 231–34.

Machery, E. (2017). *Philosophy Within Its Proper Bounds*. New York: Oxford University Press.

Maguire, E. (2022). Does memory research have a realistic future? *Trends in Cognitive Science*, 26, 1043–46.

Maguire, E.A., & Mullally, S.L. (2013). The hippocampus: A manifesto for change. *Journal of Experimental Psychology: General*, 142, 1180–89.

Martin, M.G.F. (2001). Out of the past: episodic recall as retained acquaintance. In C. Hoerl & T. McCormack (eds.), *Time and Memory: Issues in Philosophy and Psychology* (pp. 257–84). Oxford University Press.

Martin, C.B., & Deutscher, M. (1966). Remembering. *The Philosophical Review*, 75, 161–96.

Mayford, M. (2014). The search for a hippocampal engram. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369, 20130161.

Michaelian, K. (2016). *Mental Time Travel: Episodic Memory and Our Knowledge of the Personal Past.* Cambridge, MA: MIT Press.

Michaelian, K. (2021). Imagining the past reliably and unreliably: Towards a virtue theory of memory. *Synthese*, 199(3–4), 7477–507.

Moscovitch, M. (2007). Memory: Why the engram is elusive. In H.L. Roediger, Y. Dudai, and S.M. Fitzpatrick (eds.), *Science of Memory Concepts* (pp. 17–22). Oxford: Oxford University Press.

Munro, D. (2021). Imagining the actual. *Philosophers' Imprint*, 21, 17.

Nadel, L. (2007). Consolidation: The demise of the fixed trace. In H.L. Roediger, Y. Dudai, and S.M. Fitzpatrick (eds.), *Science of Memory Concepts* (pp. 177–82). Oxford: Oxford University Press.

Nelson, T.O. (1985). Ebbinghaus' contribution to the measurement of retention: Savings during relearning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 472–79.

Poeppel, D., & Idsardi, W. (2022). We don't know how the brain stores anything, let alone words. *Trends in Cognitive Science*, 26, 1054–55.

Ramirez, S., Liu, X, Lin, P., Suh, J., Pignatelli, M., Redondo, R.L., Ryan, T.J., & Tonegawa, S. (2013). Creating a false memory in the hippocampus. *Science*, 341, 388–91.

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, 426–30.

Reid, T. (2002). *Essays on the Intellectual Powers of Man*, ed. D.R. Brooks. University Park, PA: Pennsylvania State Press.

Robins, S.K. (2016). Contiguity and the causal theory of memory. *Canadian Journal of Philosophy*, 47, 1–19.

Robins, S.K. (2017). Memory Traces. In S. Bernecker and K. Michaelian (eds.), *Routledge Handbook of the Philosophy of Memory* (pp. 76–87). London: Routledge.

Robins, S. K. (2018). Memory and optogenetic intervention: Separating the engram from the ecphory. *Philosophy of Science*, 85(5), 1078–89.

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

Roediger, H.L., Dudai, Y., & Fitzpatrick, S. (eds.) (2007). *Science of Memory: Concepts*. Oxford: Oxford University Press.

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, 508–12.

Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

Schacter, D.L. (2019). Implicit memory, constructive memory, and imagining the future: A career perspective. *Perspectives on Psychological Science*, 14(2), 256–72.

Schacter, D.L., & Addis, D.R. (2007). On the constructive episodic simulation of past and future events. *Behavioral & Brain Sciences*, 30(3), 299–351.

Semon, R. (1921). *The Mneme*. London: George Allen & Unwin.

Shoemaker, S. (1970). Persons and their pasts. *American Philosophical Quarterly*, 7, 269–85.

Silva, A.J., Landreth, A., & Bickle, J. (2014). *Engineering the Next Revolution in Neuroscience*. Oxford University Press.

Stich, S. (1978). Beliefs and subdoxastic states. *Philosophy of Science*, 45, 499–518.

Sutton, J. (1998). *Philosophy and Memory Traces: Descartes to Connectionism*. Cambridge: Cambridge University Press.

Von Leyden, W. (1961). *Remembering: A Philosophical Problem*. Duckworth.

Vetere, G., Train, L.M., Moberg, S., Steadman, P.E., Restivo, L., Morrison, F.G., Ressler, K.J., Josselyn, S.A., & Frankland, P.W. (2019). Memory formation in the absence of experience. *Nature Neuroscience*, 22, 933–40.

Werning, M. (2020). Predicting the past from minimal traces. *Review of Philosophy and Psychology*, 11, 301–33.

Westfall, M. (2024). Constructing persons: On the personal-subpersonal distinction. *Philosophical Psychology*. 37, 831–860.

Wixted, J.T., Mickes, L., & Fisher, R.P. (2018). Rethinking the reliability of eyewitness memory. *Perspectives on Psychological Science*, 13, 324–35.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

Yang, W., & Yuste, R. (2017). In vivo imaging of neural activity. *Nature Methods*, 14, 349–59.

Zhang, H., Shi, Y., Jing, P., Zhan, P., Fang, Y., & Wang, F. (2020). Posttraumatic stress disorder symptoms in healthcare workers after the peak of the COVID-19 outbreak: A survey of a large tertiary care hospital in Wuhan. *Psychiatry Research*, 294, 113541.

7

Memory, Space, Time, and the Hippocampus

Charan Ranganath

Memory may be the most extensively researched topic in neuroscience, and it is arguably the greatest success story in our field.[1] Researchers have painstakingly investigated the links between synaptic plasticity, well characterized neural circuits, and large-scale brain networks in invertebrates, rodents, and human and non-human primates. These findings, which have spanned levels of analysis ranging from molecules to minds, have made it possible to propose ambitious theories to explain how processes at each level of analysis give rise to the ability to generate rich recollections of past events, to plan and navigate in the present moment, and to imagine possible futures. Overall, these developments are an indicator of scientific progress, but the sheer breadth and amount of research on learning and memory also lays bare some fundamental challenges.

There is an inherent tension between the ability to pose specific, tractable questions within the context of a particular model system and established paradigms and the ability to propose theories and make conclusions that capture findings from different model systems, paradigms, and levels of analysis. In practice, many theories of learning and memory seem to split the difference, using experimental data from a narrow range of paradigms and model systems to make sweeping claims about human cognition as a whole. Another approach has been to describe data at the neural level without resorting to theoretical constructs that describe cognition. I believe that both approaches are flawed. Instead, I argue here that we need to take on a different approach that will require both more nuance and more ambition.

My chapter, along with the others in this book, emerged from a dialogue between neuroscientists, psychologists, and philosophers at the 'Memory, Space, and Time' workshop held in Tucson, AZ from 8 to 9 November. In the first part of this chapter, I will consider how these concepts might be related to one another, and in the second part of the chapter, I will review evidence regarding the hippocampus, a brain area that is thought to support memory and the capability to orient oneself in space and time.

[1] This work was supported by a Vannevar Bush Fellowship (Office of Naval Research Grant N00014-15-1-0033) and a Multi-University Research Initiative Grant (Office of Naval Research Grant N00014-17-1-2961) from the Office of Naval Research. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the Office of Naval Research or the U.S. Department of Defense.

Charan Ranganath, *Memory, Space, Time, and the Hippocampus*. In: *Space, Time, and Memory*. Edited by: Lynn Nadel and Sara Aronowitz, Oxford University Press. DOI: 10.1093/oso/9780192882547.003.0007

You Are Here

One might reasonably ask why a group of neuroscientists, psychologists, and philosophers would want to discuss, let alone contribute to an edited volume on, space, time, and memory. From the perspective of physics, one can start from Einstein's theory of relativity to discuss the inextricable link between space and time… but that has nothing to do with memory. From the perspective of psychology and neuroscience, however, memory is the foundation that we use to meaningfully localize ourselves in space and time. I say 'meaningfully' here because I am not talking about the ability to distinguish between relatively trivial forms of localization that might be determined by looking at walls to figure out that you are in a room or looking at the position of the sun to figure out that it must be early in the morning. Instead, this chapter is concerned with the more challenging task of orienting oneself or reasoning about relative distances in space and time under conditions that cannot be directly inferred from obvious sensory cues.

Based on his observations of maze learning in rats (Tolman, Ritchie and Kalish 1946; Tolman 1948), Tolman proposed a link between memory and spatial cognition. Tolman argued that animals do not navigate in environments by learning simple stimulus-response associations, but rather that they assemble their experiences into a 'cognitive map', or a mental representation of the relationships between elements of an environment. Cognitive maps are often described by neuroscientists as a literal representation of an environment that conforms to Euclidean geometry, but that is inconsistent with Tolman's (1948) proposal. As described below, the formation, updating, and use of cognitive maps is heavily influenced by episodic and semantic memory. Neither form of memory is likely to be photographically accurate, and therefore it seems unlikely that cognitive maps will often (or ever, for that matter) conform to Euclidean geometry (Schiller et al. 2015; Epstein et al. 2017; Ekstrom and Ranganath 2018; Ormond and O'Keefe 2022; Ekstrom and Hill 2023). Instead, Tolman's conceptualization of spatial representations in his studies was more in line with the concept of a 'spatial schema' (Rinck 2005; Peer and Epstein 2021; Farzanfar et al. 2022).

I suspect that even this view is too static, and that it is unrealistic to think that we reason about space simply by retrieving information from a static schema. Instead, I believe that orienting oneself in space and time is an act of reasoning, based on a mental model (Rinck 2005) constructed through a combination of sensory information, retrieved memories for spatiotemporally dated events (episodic memory), spatial schemas, and general knowledge (semantic memory).

'Context' is the theoretical construct at the epicentre of discussions of memory, time, and space, though there are widely different conceptualizations of the term. In behavioural neuroscience, the term is often used in the most minimal, atheoretical sense, to literally refer to a box that a rodent is placed in, along with the distal cues that identify the relative location of the box in the room. Maurer and Nadel (Maurer and Nadel 2021) defined context as more of a latent variable reflecting, 'the state of the environment of the organism at any given moment, as reflected in the total

set of brain activities'. This definition emphasizes the spatial context as an animal seems to understand it. Endel Tulving's (Tulving 1972) definition of episodic memory, proposed that events occur 'at a particular spatial location and in a particular temporal relation to other events that already have occurred', or what is typically called spatiotemporal context. I agree with the idea that space and time form the context for episodic memory, however, as I noted above, the brain has no absolute measure of space or time, so our ability to orient ourselves is constructed at least in part by relying on episodic memory, which in turn is organized by context. In order to avoid chasing our tails, we need to go deeper to understand what is meant by context.

Many memory scientists, myself included, adopt a broad definition of context, encompassing factors such as physical or emotional states, goals, or any sensory information that is not at the focus of attention (Mayes, Meudell, and Pickering 1985; Ranganath 2010; Manning, Norman, and Kahana 2015; Yonelinas et al. 2019). I previously proposed that context cannot be reduced to a single feature or variable, but that we can think of context by relying on a few key principles (Ranganath 2010): (1) contexts are learned by integrating information across long timescales, (2) contexts are identified and anchored by markers that are relatively stable over time, (3) contexts are attentionally backgrounded, providing an integrated framework for understanding and directing attention to entities and objects that will be the focus of processing. As I previously noted, 'one can simultaneously process two items in the same context, but one cannot simultaneously process one item in multiple contexts' (Ranganath 2010).

In this conceptualization, context representations serve as a resource to make inferences about memory, space, and time. We can orient ourselves in space and time and judge temporal and spatial distances, in part by comparing our current context with contexts that are retrieved when we pull up past memories. A context can also serve as a cue to rapidly activate semantic representations such as a spatial schema.

Changes in one's sense of context have significant consequences for episodic, spatial, and temporal memory. This point has been made in studies of memory for simple stimuli such as word lists (Smith and Vela 2001; Manning, Norman, and Kahana 2015) but it is especially evident in memory for naturalistic stimuli such as films and stories. Zacks and colleagues have emphasized the idea that, rather than representing and experiencing our experiences as a continuum in time, people break up (or 'segment') their experiences into discrete events (Franklin et al. 2020; Zacks 2020). According to Zacks' Event Segmentation Theory, people tend to experience events as a highly predictable sequence of experiences, such as what might happen while driving down a familiar city block, watching a kick-off at a football game, or listening to the chorus of a familiar song. Events tend to be associated with inferences about goals, causes, and effects. For instance, if you were to view a single crudely drawn image from a comic strip depicting a girl kneeling and holding a football on the ground and an oddly bald boy standing nearby, you could probably infer that the girl was holding the ball so that the boy could kick it. We deploy knowledge

about such everyday events routinely, often without thinking about it, to generate an internal model of what is likely to happen.

If, however, something unpredictable happens, such as the girl pulling away the football just as the boy is about to kick it, then we must update that internal model. In Event Segmentation Theory, that moment of prediction error and model updating is called an 'event boundary', because according to the theory, at that point, one concludes that the previous event has ended, and a new one is beginning.

Event boundaries are sometimes misconceptualized as self-evident consequences of what happens in the outside world, but they are really just theoretical constructs measured by subjective reports. According to Event Segmentation Theory, event boundaries emerge as a consequence of prediction errors (information that deviates from our expectations), leading us to infer that the current event has ended and that we must form a new event model. In other words, event boundaries are internally generated based on multiple sources of information. Event boundaries can be inferred from many sources, including changes in spatial context, changes in a conversation topic, the achievement of a goal, etc. A few words could induce the sense of an event boundary, such as the phrase, 'One day later', whereas a very similar phrase like, 'One moment later', might not (Ezzyat and Davachi 2011). Likewise, a large visual change, like a shift in camera direction between two actors speaking in a scene might not elicit an event boundary, whereas a subtle shift, like one actor glancing down at their watch might. As such, event boundaries are not driven by changes in the sensory world, but rather they emerge as a natural consequence of changes in our understanding of what is happening (DuBrow et al. 2017; Franklin et al. 2020).

Event segmentation is significant for memory because people show better memory for items at an event boundary than items in the middle of an event, and they have more difficulty recalling information from a previous event (i.e., when an event boundary must be crossed) than from the current event. As noted above, spatial context changes (i.e., depicted in movies or stories, or during real-life movements) reliably trigger event boundaries, and spatial boundaries affect spatial and event memory representations (Brunec, Moscovitch, and Barense 2018; Maurer and Nadel 2021), and non-spatial context changes can affect temporal order memory for items that are separated by a boundary (DuBrow and Davachi 2013; Pu et al. 2022). These findings suggest that event boundaries reflect a change in our mental contexts, thereby shaping episodic memory retrieval, and our perceptions of temporal and spatial stability.

Hungry for the Hippocampus

The hippocampus, an area of the brain that is known to be essential for episodic memory, is widely believed to function at the crossroads of space, time, and memory (O'Keefe and Nadel 1979; Eichenbaum 2017; Ekstrom and Ranganath 2018; Sugar and Moser 2019). Research on the role of the hippocampus in memory has a long

history but there is no doubt that this work was galvanized by Brenda Milner's studies of amnesic patients like Henry Molaison (aka 'H.M.'), which sparked widespread interest in the mnemonic functions of the hippocampus. Her studies conclusively demonstrated that medial temporal lobe (MTL) damage can cause severe anterograde amnesia (Scoville and Milner 1957; Corkin 2002). Because H.M. was known to have damage to the hippocampus, the severe form of amnesia that he exhibited was often attributed to the hippocampus (despite the fact that Milner herself was more measured in her conclusions). From the 1960s through the 1970s, neuroimaging methods were not sufficiently sophisticated to characterize the anatomical extent of brain damage in amnesia patients, so it was both parsimonious and convenient for neuropsychologists to assume that all severe amnestic disorders, such as Korsakoff's syndrome, were caused by dysfunction of the hippocampus and what were often referred to as 'related MTL areas'. We now know that the hippocampus is anatomically and functionally distinct from those related MTL areas (Ranganath 2010), but there is good reason to consider how these regions interact in the service of memory-guided behaviour (Ranganath and Ritchey 2012).

I and others have reviewed the literature on the hippocampus and memory in detail (Morris 2003; Eichenbaum, Yonelinas, and Ranganath 2007; Ranganath 2010; Ekstrom and Ranganath 2018), so here, I will summarize the basic points relating the hippocampus to spatial and episodic memory encoding and retrieval (see Tables 7.1),

1. *The hippocampus is essential for episodic memory and conscious recollection in humans* (Aggleton and Brown 2006; Eichenbaum, Yonelinas, and Ranganath 2007; Mayes, Montaldi, and Migo 2007; Ranganath 2010). Patients with damage to the hippocampus are impaired at recollecting recently learned

Table 7.1 Summary of effects of focal hippocampal damage on memory performance

The hippocampus is critical for	The hippocampus is not essential for
Context fear conditioning	Cued fear conditioning
Conditioned place preference	Pavlovian conditioning or reinforcement learning
Recollection-based recognition of words, objects, or scenes	Familiarity-based word, object, or scene recognition
Temporal order memory	Conceptual or perceptual priming
Source memory	Coarse spatial memory in humans
Trace conditioning	
Context-specific extinction of cued fear	
Water maze retention (in rodents)	
Free recall	
Place recognition and object–location associative learning in rodents	
High-precision odour and object recognition	

information, and it is likely that those with hippocampal damage are unable to form rich recollections of even remote events that transpired before the brain damage occurred (Moscovitch et al. 2016). Moreover, hippocampal activity in healthy individuals increases during subjective reports of recollection (and during retrieval of objective contextual details associated with a study item), and during encoding of items that will be subsequently recollected. Notably, many aspects of memory are preserved in patients with hippocampal amnesia. Patients with focal hippocampal damage can often recognize familiar items based on a subjective sense of familiarity, and familiarity-like memory seems to be intact in non-human animals with hippocampal damage.

2. *Focal hippocampal damage impairs new spatial learning* (Ekstrom and Hill 2023; O'keefe and Nadel 1978; Kesner 2003; Banta Lavenex and Lavenex 2009; Epstein et al. 2017). Hippocampal damage in rodents impairs learning in a wide range of spatial memory tasks including the Morris Water Maze, Barnes maze, and the radial arm maze, and it reduces the acquisition of indirect expressions of spatial memory, such as context fear conditioning. Hippocampal damage in humans slows spatial learning (Kessels et al. 2001) and reduces spatial memory precision (Kolarik et al. 2016), and anecdotally, hippocampal amnesic patients seem lost in space and time, but they do not exhibit extreme spatial disorientation (e.g., bumping into walls) (Aguirre and D'Esposito 1999).
3. *The hippocampus seems to be critical for forming integrated memories across overlapping events* (Eichenbaum et al. 1999; Shohamy and Wagner 2009; Schlichting and Preston 2015; Schapiro et al. 2017; Bellmund et al. 2018). For instance, the hippocampus seems to be critical for transitive inference tasks (Dusek and Eichenbaum 1997), one must learn to integrate relationships between items across multiple learning trials (e.g., if A > B, and B>C, then A>C), and hippocampal activity reflects inferred relationships in related paradigms in which transitive relationships must be integrated across two different dimensions (Park et al. 2020). The hippocampus also contributes to statistical learning and sequence learning tasks that require learning of relations between temporally adjacent items and associative inference tasks in which participants learn to infer relationships between information learned in overlapping associations. One caveat, however, is that it is not clear whether the contribution of the hippocampus is primarily due to fast encoding of individual memories (which is necessary in order to do integration tasks) or if it makes a special contribution to memory integration per se.
4. *Neurons in the hippocampal system (i.e., EC, DG, CA1, CA2, CA3, Subiculum) carry information about spatial, temporal, and task context* (Cohn-Sheehy and Ranganath 2017; Ranganath 2019; Ranganath and Ekstrom 2020; Antony et al. 2023; Eichenbaum and Cohen 2014; Sugar and Moser 2019). The most well-studied examples are hippocampal place cells, which show activity when an animal is in a particular location in an environment, and entorhinal grid

cells, which have multiple place fields arranged in a hexagonal grid. The hippocampus is also known to have 'time cells' that fire in a sequential manner, tracking even empty temporal intervals when an animal is in one place. In both the lateral entorhinal cortex (LEC) and in hippocampal subfields CA1 and CA2, neurons exhibit 'temporal drift', meaning that cell firing and/or selectivity changes over time. Within LEC, different subsets of neurons seem to fire at different intervals during recording sessions, with some abruptly increasing and then ramping down, and others ramping up to a peak and then abruptly dropping (Tsao et al. 2018; Umbach et al. 2020). In both humans and rats, the population-level dynamics in LEC showed a 'drift' over time, with different subsets of neurons coming on and offline at different intervals, much like what has been proposed as a basis for representing temporal context in memory. Likewise, populations of hippocampal place cells have been shown to drift over time, such that the same location might be encoded by different populations of neurons at different times.

5. *Population-level hippocampal activity in humans during episodic memory tasks reflects the relationship between specific items and the context that they were encountered in* (Ranganath 2010; Ranganath and Ekstrom 2020). With fMRI, one can use analyses of activity patterns across different hippocampal voxels to index information carried during memory retrieval (Dimsdale-Zucker and Ranganath 2019), much like neurophysiologists can analyse the patterns of calcium activity or spiking in large populations of neurons. Both approaches parallel how one might infer the presence of a representation in a neural network, in the sense that they reflect how the population dynamics relate to different experimental manipulations. In virtually every study that has attempted to pull apart hippocampal representations of items (e.g., words, faces, or objects) and contexts (e.g., a spatial context, the temporal context provided by a structured list of items, a movie, or a temporal sequence, or a particular orienting task used to help the subject encode an object), hippocampal representations carry information about items in context. For instance, hippocampal activity patterns might reliably index memory for a particular object in a temporal sequence and yet carry little to no information about the same object in a different sequence context (Hsieh et al. 2014).
6. *Hippocampal activity is modulated by predictability and salience* (Ranganath and Rainer 2003; Lisman and Grace 2005; Lisman, Grace, and Duzel 2011). This is perhaps the most widely known, and yet most poorly appreciated aspect of hippocampal function. Population-level activity in the hippocampus enhances dramatically during a surprising, novel (e.g., viewing a novel scene), or motivationally significant (i.e., cues that signal or direct responses to reward, pain, threat, mating possibilities, etc.) event. Such experiences are also associated with enhanced neuromodulatory release and activity in numerous brain areas, so it might be tempting to dismiss these results as uninformative when it comes to hippocampal function.

More surprising, however, is that hippocampal responses to salient or novel experiences are effective biomarkers of hippocampal, and episodic memory functions (Grunwald et al. 1999; Billette et al. 2022). For instance, hippocampal field potentials elicited by novel or surprising stimuli can be used to determine the integrity of the hippocampus to guide surgical interventions for individuals with severe epilepsy. One study (Grunwald et al. 1999) even found that the magnitude of hippocampal field potentials to surprising stimuli strongly predicted cell counts in Dentate Gyrus (DG) subfield. Intracranial recordings in humans have also revealed neural correlates of successful encoding and retrieval in basic memory paradigms, but the magnitudes of such effects are relatively weak by comparison to novelty and surprise effects and they are not evident in every subject even when the hippocampus is functional (Axmacher et al. 2010).

7. *Hippocampal activity during complex events emphasizes spatial and event boundaries* (Brunec, Moscovitch, and Barense 2018; Franklin et al. 2020). Activity in hippocampal place and grid cells is determined by environmental boundaries, and crossing of boundaries between compartments triggers remapping (i.e., an abrupt change in the population-level spatial representation) of place cells (Spiers et al. 2015). The presence of hippocampal boundaries modulates activity during imagination of scenes (Bird et al. 2010), and during memory retrieval, the hippocampus preferentially represents previously encountered objects in proximity to a major spatial boundary (Geva-Sagiv et al. 2023). Event boundaries in films and stories also elicit increases in hippocampal activity measured with fMRI (Baldassano et al. 2017; Ben-Yakov and Henson 2018; Reagh et al. 2020; Cohn-Sheehy et al. 2021; Barnett et al. 2023) and single-unit recording (Zheng et al. 2022). Hippocampal activity increases at event boundaries (at least in films) have been reliably demonstrated in numerous studies, and in a study of over 540 subjects, we found that such increases reliably predict individual differences in event memory outside of the scanner (Reagh et al. 2020). We also recently found that, when repeatedly viewing a movie, the hippocampus carries information about the movie experienced during the boundaries but not from the middle of the event (Reagh and Ranganath 2023).

   Lu and Norman (Lu, Hasson, and Norman 2022) ran computational simulations suggesting that, if we segment our experiences into events, event boundaries may be optimal points in time to encode the events into memory. Consistent with their proposal, we found that increases in hippocampal activity and functional connectivity between the hippocampus and neocortical networks that encode event information predict successful encoding of the entire preceding event (Barnett et al. 2023). Finally, the hippocampus also appears to contribute to memory integration at event boundaries such that past and present information can be linked into a seamless narrative (Cohn-Sheehy et al. 2021).

Theories of Hippocampal Function in Relation to Episodic and Spatial Memory

The hippocampus has a large fanbase, and one could generate an entire series of books on theories of hippocampal function. Most of the theories fall into three broad classes,

Cognitive Map-Based Theories

As noted above, Edward Tolman (Tolman 1948) proposed that animals represent their experiences via a cognitive map that enables one to use memory in a flexible and integrative manner. For instance, if I were to drive to the supermarket and discover that the road that I usually take is closed, I could figure out a number of alternate routes to reach my destination, including even a route that I have never taken before. If I had only learned to navigate via simple stimulus-response associations (i.e., statistical learning), this would not be possible, as I could only retrieve a chain of pre-experienced events.

In an ambitious synopsis of a large range of findings, O'Keefe and Nadel (1978) extended Tolman's concept of a cognitive map to explain the role of the hippocampus in memory. Their monograph included an expansive review of philosophical frameworks for spatial cognition, spatial navigation behaviour in different model systems, lesion and physiology data linking the hippocampus to spatial navigation and memory, research on human amnesia, and even the role of spatial representations in linguistic representation. Interested readers should read O'Keefe and Nadel (1979), which presents a focused and concise summary of their argument, along with responses to a number of critiques made by others at the time.

O'Keefe and Nadel's (1978, 1979) thesis was predicated on the idea that the vertebrate hippocampus evolved to solve the problem of spatial orientation and navigation. Building on Endel Tulving's definition of episodic memory as a system that is organized by spatiotemporal context (Tulving 1972), they proposed that the hippocampus also supports episodic memory. In addition, they proposed that, in humans, the right hippocampus supports semantic memory, by providing a neural framework for relating concepts within a feature space.

O'Keefe and Nadel proposed that the hippocampus builds representations that exhibit the following properties: '(1) preservation of spatio-temporal context; (2) single occurrence storage; (3) minimal interference between different representations of the same item; (4) multiple channels of access for the retrieval of any, or all, of the relationships embodied in the map'.

In the years since it was first proposed, O'Keefe's conception of Cognitive Map Theory (CMT) became increasingly focused on spatial relationships (Hartley et al. 2014), whereas Nadel and Moscovitch (Nadel and Moscovitch 1997;

Moscovitch et al. 2016) emphasized a role for the hippocampus in representation of spatiotemporal context in the service of episodic memory. Other researchers have since doubled down on O'Keefe and Nadel's speculations about representations of language and verbal memory in a spatial framework, proposing abstract and far-reaching models suggesting that the hippocampus can represent any kind of relationships in memory (Schiller et al. 2015; Behrens et al. 2018).

The most well-known model in this vein is the relational memory theory (RMT) of Neal Cohen and Howard Eichenbaum (Cohen and Eichenbaum 1993; Eichenbaum and Cohen 2001, 1993). O'Keefe and Nadel (1978) explained human memory through an evolutionary lens, such that episodic memory could be seen as a relatively simple extension of a neural system that evolved in vertebrate animals to support spatial navigation. Cohen and Eichenbaum, in contrast, started with on the premise that any theory of memory and hippocampal function should be focused to explain the deficits and intact abilities seen in patients with amnesia. At that time, a great deal of research had shown that patients with severe amnesia could still express memory indirectly, without conscious awareness (Cohen and Squire 1980). Cohen and Eichenbaum proposed that human amnesia and research on spatial processing in rodents provided converging evidence for the idea that the hippocampus represents relationships between items in memory, or 'declarative memory' (Cohen and Eichenbaum 1993).

RMT and CMT are often viewed as competing theories, but a careful reader will note that the two theories are largely aimed at explaining the same findings with the same general set of ideas—that the hippocampus encodes relationships between elements in memory, and in humans, this system is capable of encoding relationships that extend beyond physical space. Eichenbaum (Eichenbaum et al. 1999; Eichenbaum 2017) made this point more explicit when he argued that the hippocampus encodes a 'memory space' that relates experiences to one another. That said, there are a few notable differences between CMT and RMT. Like O'Keefe and Nadel, Eichenbaum noted that spatial relationships are a significant variable that must be encoded relationally, but he argued that other important variables, such as olfactory cues, are of equal evolutionary significance. Thus, RMT did not assert that spatiotemporal context is the primary organizing dimension in the hippocampal 'memory space'. Additionally, Cohen and Eichenbaum rejected the view that the hippocampus encodes space in a Euclidean manner, instead proposing that representations are simply relational. This would correspond to topological coding of a physical space, such that an object could be seen as 'to the left of' or 'close to' a particular landmark, rather than encoding an object with precise distances and angles. That said, O'Keefe and Nadel acknowledged that the hippocampal system might have changed with the evolution of language, enabling it to support more abstract relationships. Viewed from a distance, I see a number of inconsistencies in the various descriptions of RMT and CMT, and as far as I can tell, the primary point of disagreement between these two viewpoints is whether the theory should start from

explaining hippocampal function in human amnesia or whether hippocampal function should be best explained under the broad umbrella of evolutionary demands met by all vertebrate animals.

Somewhat ironically, the awarding of the Nobel Prize for the discovery of hippocampal place cells and entorhinal grid cells inspired a resurgence in more abstract models of a hippocampal memory space (Schiller et al. 2015; Behrens et al. 2018; Bellmund et al. 2018). These models inspired numerous studies examining brain activity during tasks with variables that differ along two continuous dimensions. The general finding is that hippocampal and EC activity seems to reflect relationships (i.e., distances or angles) in any kind of 2D space (Bottini and Doeller 2020). David Tank and colleagues broadened this view a bit (Nieh et al. 2021), arguing that the hippocampus maps any experience on a low dimensional neural manifold (i.e., a representational space that reflects the possible trajectories of neural activity in a neural network over time). It is not clear how to interpret the tasks described above, however, because a task with relevant variables that vary only along two dimensions does not provide a strong opportunity to identify whether the hippocampus might ever encode information that do not conform to a map-like structure.

Later, I will discuss some limitations that are common to many theories, but here I will simply note that the flexibility of RMT and similar memory space theories to explain a broad range of data from non-spatial learning paradigms is also a potential weakness. If the hippocampus can encode relationships between anything and everything, this would create a combinatorial explosion. The world can be carved up in infinite dimensions, and thus the idea that the hippocampus can reduce any experience into a low-dimensional manifold cannot make sense … that is, unless the challenging task of determining the relevant dimensions at any given moment is outsourced to neocortical regions that actually do the heavy lifting (Ekstrom and Ranganath 2018; Ekstrom and Hill 2023).

Memory Binding Theories

CMT and related theories focus primarily on the hippocampus, but implicit in these theories, and to my knowledge, every theory of hippocampal memory functions, is the idea that the hippocampus encodes a memory 'index' (Teyler and DiScenna 1986; Teyler and Rudy 2007). The indexing theory assumes that memory retrieval involves activating the same (more or less) neocortical cell assemblies that were activated during a past event. Given that the hippocampus receives inputs from diverse areas of the brain during conscious experiences, it is reasonable to think that the hippocampus rapidly encodes memories that capture the pattern of neocortical activity at a given moment. Later, if that hippocampal memory trace is reactivated, the brain can essentially work backwards to reinstate the patterns of activity associated with that past event.

Mishkin et al. (Mishkin et al. 1997) elaborated on the indexing theory to accommodate how neocortical areas that provide the majority of the input to the hippocampus—the perirhinal and parahippocampal cortex (PRC and PHC)—might support memory independently of, and in concert with the hippocampus. Their model proposed a hierarchy of mnemonic representations in the MTL, such that the PRC encodes memories for objects and the PHC encodes memories for spatial contexts, and the hippocampus integrates inputs from these areas, associating (aka *binding*) the 'what' and 'where' inputs into a coherent representation. According to their model, these neocortical areas may be sufficient to support semantic memory (though the relationship between what and where systems was rather murky), and the hippocampus, being uniquely situated to bind object and spatial context representations, is necessary to support context-rich episodic memories. Since its publication, the anatomical framework proposed by Mishkin et al. (1997) has been challenged by more detailed neuroanatomical findings which complicate the simple idea of 'what' and 'where' streams that converge in the hippocampus (Kravitz et al. 2011; Nilssen et al. 2019), but the basic ideas in that framework remain influential.

The hierarchical model of MTL memory functions was soon expanded to account for the flexibility of human memory. According to one class of models, which I have referred to collectively as the 'Binding of Items and Contexts' (BIC) model, the perirhinal cortex encodes memories for *items*, encompassing not only objects but also words, faces, or even abstract concepts about these stimuli, and the parahippocampal cortex encodes memories for *contexts*, including scenes and also non-spatial elements that are processed in the background and remain stable over time (Eacott and Gaffan 2005; Davachi 2006; Knierim, Lee, and Hargreaves 2006; Eichenbaum, Yonelinas, and Ranganath 2007; Mayes, Montaldi, and Migo 2007). In this more abstract framework, the hippocampus forms memories that *bind item and context* information (Diana, Yonelinas, and Ranganath 2007), such that being in a particular place, or smelling a particular odour, can bring back a wealth of other disparate elements that are bound together solely because they occurred at a particular place and time. This recollective process need not depend on any intrinsic circuitry in the hippocampus, as it could emerge simply because hippocampal neurons are forming arbitrary connections with neocortical areas that encode the basic components of an episodic experience. In this way, BIC builds on the basic premise in RMT, in that the job of the hippocampus is to perform all-purpose 'relational binding'.

In contrast to RMT and CMT, BIC (and related frameworks) explains hippocampal contributions to human episodic memory as well as forms of explicit memory that are *not* supported by the hippocampus. Aggleton and Brown, for instance, did not emphasize binding, item, or contextual information per se, but they built on Mishkin's idea proposed that the hippocampus supports recall and conscious recollection of past experiences, and suggested that the PRC is sufficient to support item recognition based on familiarity (O'keefe and Nadel 1978; Aggleton and Brown

1999; Mayes and Montaldi 2001). Later, Howard Eichenbaum, Andy Yonelinas, and I amassed a range of converging evidence from fMRI, amnesia, and animal lesion studies expanding on this point by showing that recollection of both spatial and non-spatial contextual information also engages the PHC (Eichenbaum, Yonelinas, and Ranganath 2007). Notably, revised versions of the BIC framework are compatible with the version of CMT proposed in O'Keefe and Nadel (1978), suggesting that the hippocampus uniquely supports event-specific memory because hippocampal representations are fundamentally organized along a dimension of spatiotemporal context (Ranganath 2019; Ranganath and Ekstrom 2020).

What Can We Learn from Computational Models of Memory?

Countless papers describe the 'computations' implemented in the hippocampus, but it is sometimes hard to know whether such phrases have any explanatory power. That said, computational models of learning can provide some insights into the functions of the hippocampus, and how they might relate to the bigger questions about space, time, and memory.

Given the ubiquity of synaptic plasticity, we can generally assume that any network of interconnected neurons can support a kind of learning, and neural network models provide a crude, but useful way to understand learning processes. At the most basic level, a neural network is a device that learns to lump and split, and any particular learning process can ultimately be reduced to the problem of when to lump and when to split. That is, a model encodes a series of input patterns and learns (either on its own or via explicit instruction) to assign the inputs to a latent representation of the input. If your goal is to categorize or classify inputs, such as differentiating between syllables like 'BA' and 'PA', then you want a model that takes a bunch of different inputs and lumps them into a few representations. If your goal is to remember singular experiences, as in episodic memory, then you want a model that splits, creating a new representation for anything that is slightly different from what has been previously learned.

The pioneering computational neuroscientist David Marr pointed out that the anatomical characteristics of networks in the brain determine the network's propensity to lump or split. In the span of only two years, Marr proposed computational frameworks for understanding the functions of the cerebellum, neocortex, and hippocampus based on their neuroanatomical characteristics (Marr 1969, 1970, 1971). It is hard to overestimate the importance of these papers for all branches of modern neuroscience, but here I will focus on Marr's most significant proposal, that the 'neocortex is capable both of classifying and of memorizing inputs', whereas the hippocampus, 'is capable only of memorizing them'.

In essence Marr proposed that the neocortex is a lumper, assigning inputs to 'memories' that capture the most meaningful elements of that input. By this view, the neocortex does not literally memorize inputs, but rather it encodes an 'internal

description of the environment … which constitutes the animal's memory of the information'. Just as a bird expert can exploit their expert knowledge to quickly learn information about new birds that they see for the first time, a neocortical network that has picked up the structure of past experiences can rapidly encode new memories by classifying them in relation to other previously encoded memories. In the neocortex, everything is connected.

Simple connectionist models, complex deep neural networks, and even the most advanced tools of generative artificial intelligence (AI) with billions of parameters epitomize this idea—they are not designed to encode singular experiences, but rather to capture structural similarities across multiple inputs (Hassabis et al. 2017). Put another way, 'lumper' networks learn the structure of the data that they are trained with, so that they can efficiently generalize what was learned to new examples. For instance, if a network were to learn the attributes of a series of birds (crow, eagle, sparrow, etc.), with each learning experience, the weights in the network would be successively tweaked to capture that category structure, optimizing the ability to learn about new birds and make inferences about birds that it has never encountered.

Marr recognized that the strength of neocortical networks at learning structure is also a weakness, in that they are poorly suited to learn information that violates that structure. For instance, a cortical network that has learned the canonical features of birds would have trouble encoding information about a penguin—a bird that swims but does not fly. He argued that the hippocampus helps the neocortex to overcome this problem by implementing a separate form of learning called, 'simple memory', which involved memorizing specific experiences, as opposed to discovering relationships between different experiences. In other words, Marr felt that the hippocampus is designed to be a splitter, unable to categorize or classify experiences, and instead storing 'simple' memories that are unrelated to one another. Marr speculated that it could be useful to have a system that can quickly encode individual events (splitting), so that these simple could be used by the neocortex to identify statistical patterns that amongst these experiences (lumping).

Marr's intuition about the computational trade-offs between learning the statistical structure of training data and the ability to rapidly and flexibly encode new information incompatible with that structure has been elaborated upon by other investigators (Grossberg 1980; McCloskey and Cohen 1989; McClelland, McNaughton, and O'Reilly 1995). Grossberg (1980) described this trade-off as the 'Stability-Plasticity Dilemma', and McCloskey and Cohen (1989) demonstrated how, if the same neural network were trained to rapidly learn new information that is incompatible with previously learned information, then the model will suffer from 'catastrophic interference', and lose the previously learned information. McClelland, McNaughton, and O'Reilly (1995) argued that the hippocampus helps the brain deal with this trade-off, such that it is able to enable rapid encoding (i.e., 'simple memory' in Marr's terminology) of arbitrary information, so that the neocortex can slowly

incorporate the new information with previously acquired knowledge. This division of labour bears a great deal of resemblance to the distinction between semantic and episodic memory, which as described earlier, are differentially dependent on the hippocampus and neocortex.

To make this more concrete, imagine a commuter who must park their car in a crowded neighbourhood every morning. Over time, this commuter comes upon a street where she is likely to get a spot, but one morning she gets an email saying that the street will be temporarily closed for construction. Our commuter's brain has a dilemma because she needs to remember that she can't count on her usual spot on that day, but because the closure is temporary, she does not want to unlearn the location of her favourite parking area. In other words, she needs a neocortical semantic-like memory network that lumps in order to figure out the general locations where parking is likely, and she needs a hippocampus episodic-like memory network that splits so that she can rapidly learn that the street is temporarily closed.

As it turns out, the hippocampal system has two computational tricks that make it ideally suited for rapidly encoding new experiences in a manner that keeps them separated from other similar experiences (Hasselmo and McClelland 1999). One mechanism for overcoming interference between different events is inhibition, which is very high in the EC[2] and DG. As a result, the overall magnitude of activity in the hippocampal system is very low (or 'sparse'), and particularly in DG (O'Reilly and McClelland 1994). (As an aside, this makes single-unit recordings in DG very challenging because the overall level of activity is so sparse that it is hard to get sufficient data to make strong conclusions.) Inhibition is important because it determines the signal to noise ratio of a network, and in the most extreme case, it can even serve as a gate to prevent encoding of weak inputs. Consider what happens if inhibition were low: two memories that have common features that have overlapping neocortical representations would activate overlapping populations of neurons in Entorhinal cortex layer 2 (ECII). But when inhibition is high, then an input pattern from the neocortex would activate only a small group of neurons that are most strongly activated by the inputs. Thus, with very high inhibition, activity in the network is sparse, but the neurons that are active are highly diagnostic of the particular input.

The second mechanism for mitigating interference is binding. Binding happens because inputs from diverse cortical pathways converge onto single neurons in the EC and DG (Nilssen et al. 2019). If the different cortical inputs to the hippocampus each convey information about different elements of an event, hippocampal binding should help to disambiguate memories that have overlapping elements. As an analogy, a Thai curry recipe and a pina colada cocktail recipe could both require a can of coconut milk, but despite the overlap in the ingredients across the two recipes, once you combine the coconut milk with the other ingredients, the end results for the two

[2] Here, I am particularly emphasizing layer IIa of the EC and the DG.

recipes would taste very different. In the same vein, if a population of neurons were to randomly conjoin information about features of people and things in an event with information about the spatiotemporal context of the event, then even memories with the same elements would become differentiated from one another. Thanks to high inhibition and binding, two different events that activate overlapping neural populations will be squashed, blended, and compressed to the point that they map onto distinct clumps of DG neurons. This process is called 'pattern separation' (Wigström 1973).

Pattern separation is, in my mind, the most computationally meaningful operation that occurs in the hippocampal system, but I will not dwell on it here. The key point is that the hippocampus is biased to be a splitter, not a lumper. This goes some way to explain its critical role in encoding precise memories for events that occurred in a particular spatiotemporal context, because this requires highly distinct neural event representations. This is not to say that the hippocampus cannot lump, however, nor does it mean that the neocortex cannot split. Relevant to the former point, in studies of memory integration, the hippocampus might retrieve a previous association when a new overlapping association is encountered. If that memory is updated to accommodate the new information, then the hippocampal representation will now be less distinctly tied to any one experience (Schapiro et al. 2017). In other words, the form of neural networks shapes the function, but it is not a very strong constraint.

It's Complicated

I've tried very hard to provide a succinct review of empirical evidence and theories about the role of the hippocampus in representation of memory, space, and time, but I am sure any reader will conclude that there is simply a mountain of work on these topics. And I have skipped over a number of relevant topics and ideas. How do we deal with the complexity and scope of what has been found?

One might simply focus on explaining a narrow range of data. Some degree of reductionism is necessary for the scientific enterprise, and it makes sense to focus on narrow questions that are posed in a manner that will yield specific answers. For instance, we could focus on the ubiquity of spatial modulation in hippocampal neurons and embrace a 'space first' variant of CMT which proposes that the hippocampus was designed to map physical space. We could fit all sorts of models to data explaining how place cells and grid cells systematically map spaces within small, enclosed environments, but we might have to avoid thinking about the fact that people with hippocampal amnesia don't bump into walls. We would also probably have to ignore the instability of hippocampal population codes in response to minor contextual changes and assume that somehow the brain is smart enough to ignore the neurons that remap and only use the population of neurons that retain their selectivity as a stable cognitive map.

Alternatively, we can adopt the view that the hippocampus is an all-purpose memory space, relating all of our experiences to one another as points in a low-dimensional space. I could show how (some of) the same principles that apply to place and grid cells during physical navigation have compelling parallels to single-unit activity in rodents and fMRI results in humans performing tasks that have stimuli or task relevant variables that follow a two-dimensional structure. If we go this route, we will have to ignore the findings that are problematic for the 'space first' view, and we will have to avoid thinking about whether someone with hippocampal damage is capable of performing a simple task that required moving a joystick to change a sound or visual stimulus.

Alternatively, we can focus on the fact that people with hippocampal damage have problems with episodic memory, and that measures of hippocampal activity are sensitive to successful episodic memory retrieval are accompanied by information about the spatiotemporal context of retrieved events. In this case, we could embrace the BIC model, in which place cell coding is explained as part of a spatiotemporal code that we use to encode unique episodes. But we will also need to ignore data showing that hippocampal activity patterns seem to carry information about spatial distances between locations in physical and virtual spaces, independent of temporal distance. And we will have to ignore the fact that place cells seem to be pretty stable even across multiple episodes within the same environment.

As a last resort, we could take a bird's-eye view and say that all of the theories that I have described are simply blind men groping different parts of a seahorse. We could say that these poor blind men are also aphasic, as their words are meaningless. That is, words like episodic memory, or the words that we use to describe our conceptions of space and time, and pretty much every other word is just an emergent property of chaotic activity patterns in the hippocampus. Having played the 'dynamical system' card, we could simply absolve ourselves of the responsibility to explain anything, protecting us from the possibility of ever being wrong. Unfortunately, we would also have to ignore the existence of children who might benefit from accommodations due to learning disabilities, soldiers with traumatic brain injuries, older adults who might benefit from interventions to prevent neurodegenerative diseases, stroke patients who might be incapable of living independently, and so on. When we abdicate responsibility for explanation, we abandon the opportunity to help people make informed choices that will have far reaching consequences on their lives. If we neglect behaviour, cognition, and theory, we will render ourselves useless.

The Middle Path

According to the Buddha, 'There is a middle way between the extremes of indulgence and self-denial, free from sorrow and suffering'. In our case, I think there is also a middle way between the extremes of overfitting to a narrow range of evidence and denying ourselves the opportunity to explain anything (Ranganath 2022). As a

bonus, I think we can even make the fruits of our endeavours a bit more useful. Let's consider three points that might get us in the right direction.

1. *Words are important.* It is true that computations in a dynamical system cannot be reduced to a simple verbalizable concept, but most everyone—even computational neuroscientists—uses words like 'space', 'time', and 'memory', as well as more interesting words like, 'navigation', 'imagination', 'simulation', and so on. We use these words to communicate the importance of our research, but it is necessary to think about whether our methods enable us to relate our findings to words as understood by other humans. We often rely on narrowly framed experimental questions to reduce complexity and get a clear answer, but then we use words to summarize conclusions from these findings. The key question is whether our experiments have a valid relation to the concepts that we are attempting to study.
2. *We must consider the limitations of model systems and paradigms.* A house built on place cells cannot stand. Model systems and experimental paradigms are necessary in order to pose questions and generate useful data, but we must also deal with the fact that our model systems and paradigms fundamentally shape our data, and by extension, they constrain the kinds of answers that we can get. Here are two examples.

   As I have written elsewhere, we have learned an awful lot about the brain from the random foraging paradigm used in most studies of place cells, but there are limits (Zucker and Ranganath 2015; Ekstrom and Ranganath 2018). These studies typically involve a rodent (usually rats, but sometimes mice or bats) that has little knowledge about the world. After it has had time to recover from brain surgery, it is grabbed by a human, hooked up to some strange devices, and placed in a box where it chases food pellets that are tossed in plain sight. I can't imagine how rats experience this, but I imagine it is what an infant would feel like if they were thrown into the middle of an IKEA showroom.

   Rats, unlike humans, do not have foveal, stereoscopic vision and they rely heavily on olfaction and whisking to learn about new environments, so they orient to distal cues and must move throughout the box to develop an internal model of the space. As the rat chases food pellets in that meticulously scrubbed box, there is not much else of importance beyond the context, and the scattering of food pellets ensure that every part of the context is significant. Unsurprisingly, these are the best conditions to identify hippocampal neurons that show pure spatial selectivity. However, when cues are manipulated in virtual reality, one can find that the hippocampus preferentially encodes whatever information is salient and useful (Moore et al. 2021). Moreover, when an animal is moving in order to do a task (like spatially navigating to a goal location), hippocampal neurons encode the task relevant information. And finally, in primates that explore the world through saccadic eye movements (Shen et al. 2016), hippocampal neurons often reflect oculomotor variables (as well

as information about sensory stimuli and task variables). All this is to say that we cannot assume that a finding recorded from a simple paradigm like random foraging has implications for hippocampal function in other paradigms or across species.

Let's now consider an example from human research. I have spent much of my own career studying memory in humans by having them memorize lists of words or pictures that are briefly presented in a random (or pseudo-random) order. List learning is a valuable experimental paradigm, but it also has limitations. The subject typically remains still during these paradigms, passively viewing or making decisions about items that flash briefly at the centre of a screen, and if the experiment is designed properly, each stimulus should be unpredictable. Under these conditions, even college students (who are extremely skilled at rote memorization and test-taking) do not perform perfectly. In list learning paradigms, we see that hippocampal activity is higher for items that are forgotten than items that are recollected during the test phase. And patients with hippocampal damage (or for that matter, middle-aged and older adults, children, those with psychiatric or neurological disorders, etc.) will typically show impaired performance relative to 'healthy controls'.

Does this mean that the hippocampus 'does' episodic memory? To answer this question, we must understand what is measured in list learning paradigms. Is each item an 'episode'? Or is the entire list an episode and each item is a detail that is part of the episode? I would bet that if you were to call back a subject from a list learning experiment and ask them to recall events from their day in the lab, they would describe sitting at a computer, seeing a bunch of words, and pushing buttons to make decisions on a test. In other words, the most significant elements of the episode are not directly measured in a list learning paradigm (Ranganath 2022).

3. *Neural interactions are important.* I never formally trained in neuroscience, and when I started out as a neuropsychologist, I thought of the brain as a series of specialized modules that do their job separately. In the neocortex, we could see that damage to discrete areas could cause neglect, aphasia, apraxia, agnosia, or amnesia depending on the site of the lesion. In the case of memory, I thought of the hippocampus as the memory store, and the prefrontal cortex as the executive, controlling the use of memory in the service of action. I am not the only one who has thought of the brain in this way, and although it is not a bad starting point, we need to move on.

   Different areas of the brain are not independent of one another. What happens in one area is going to have consequential impact on activity in other areas that are anatomically connected. Anyone who has a basic understanding of statistics should understand that an interaction is different from two main effects, and yet most theories in neuroscience (particularly those involving the hippocampus) seem to assume that information is serially processed

in the brain, with information handed off from one area to another in a series of steps that ultimately leads to behaviour.

If we accept that networks are important, then we must consider the kinds of interactions that might be relevant to the topics of memory, space, and time. There seems to be a general consensus to the idea that complicated activities like episodic memory retrieval or spatial navigation depend on networks that involve neocortical areas and the hippocampus, and we would assume that learning about spatial contexts and events should be associated with plasticity in these networks. Finally, I think most would agree that episodic memory and spatial navigation are not equivalent, despite relying on many of the same brain regions. Accordingly, there must be some differences in the brain areas that are involved, and in the way brain areas interact with each other over the course of these activities. Thus, the hippocampus and neocortex are not just complementary learning systems, they are interacting memory systems, and therefore the 'function' of the hippocampus will be inextricably linked to its interactions with a broader network context, which in turn should depend on the behavioural context.

Following this line of argument to its logical end, it becomes apparent that we should abandon the idea that we can arrive at an all-encompassing explanation of the functions of the hippocampus. The same limitations apply to a simple 'brain inside-out' (Buzsaki 2021) approach (in which we focus on what emerges from neural activity rather than focusing on function), because our descriptions of the working of neural circuits in regions like the hippocampus will only be interpretable in the context of what is happening in other key regions that are interacting with the hippocampus at the moment. Put another way, our observations of neural activity cannot be separated from the context in which the activity is observed, and the way it is measured.

Where Do We Go from Here?

A broad strokes view of memory, space, or time is unlikely to yield fruitful progress. If we want to ask if the hippocampus is part of the brain's circuitry for spatial navigation, or whether the hippocampus is an all-purpose memory device, mapping relationships between any and all kinds of experiences, the answer will be 'it depends'. If we consider that memories are constructed through dynamic interactions between the hippocampus and particular neocortical networks, and that the interactions can dramatically change based on motivations, goals, context, and prior knowledge, then we cannot assign a single function to the hippocampus.

And even if we broaden out our scope beyond the hippocampus, and we focus instead on cognitive constructs, it is not hard to see that 'space' or 'time' isn't the right target for scientific inquiry. Our sense of where we are is related to, but not equivalent to, spatial navigation. Our sense of orientation in time is related to, but

not the same as the ability to estimate temporal durations or the ability to remember the temporal order of past events. Our ability to recognize faces or words is related to, but not the same as reconstructing an episodic memory for a complex event.

So, what is the best way forward? I think there is a useful analogy to be made with genetics research. Most common diseases cannot be explained by single genetic polymorphism, and the field has likewise failed to explain inter-individual variability in behaviour based on the simple effect of a single gene or even a main effect of multiple genes. This is because genes interact with one another and because they can only be understood in the context of environmental influences (broadly defined). That said, valuable progress can be made by studying genetic influences within a relevant context (such as studies of the interaction between genetic risk and adolescent marijuana use in the development of schizophrenia, Karl and Arnold 2017).

We can take a similar approach to ask questions about the brain and behaviour. We need to focus on specific problems and rely on evidence that comes from paradigms that are directly relevant to that problem. For instance, we can develop models to explain how we orient ourselves in space, using behavioural paradigms and concurrent neural measures that directly get at the problem of spatial orientation (as opposed to visually guided movement). We would need to also be careful not to conflate orientation with navigation, which occurs in the context of plans and goals. And rather than assuming a precise, photographic representation of space, we would need to understand the influence of prior knowledge.

We can also develop models to explain how we recall complex events that occurred at a particular place and time and collect neural data in experimental paradigms that capture how real world events are understood and represented. Again, useful information can be derived from human and animal models, provided that one accounts for the important role of prior knowledge.

In other words, we need to be sceptical about efforts to build a grand theory of everything (much to the chagrin of anyone hoping for the universal algorithm for cognition), and we need to think carefully about how our paradigms, model systems, methodological tools, and theories can be optimized to target more focused questions. In other words, if there is a middle path, it might require us to be more ambitious in our approaches and yet more pragmatic in acknowledging the complexity of extrapolating from laboratory research real-world phenomena.

References

Aggleton J.P., and Brown M.W. (1999). Episodic memory, amnesia, and the hippocampal-anterior thalamic axis. *Behavioral and Brain Sciences* 22, 425–44.

Aggleton J.P., and Brown M.W. (2006). Interleaving brain systems for episodic and recognition memory. *Trends Cogn Sci*, 10, 455–63.

Aguirre G.K., and D'Esposito M. (1999). Topographical disorientation: A synthesis and taxonomy. *Brain*, 122, 1613–28.

Antony J., Liu X.L., Zheng Y. et al. (2023). Memory out of context: Spacing effects and decontextualization in a computational model of the medial temporal lobe. *Psychological Review*, 131(6), 1337–1372. doi: 10.1037/rev0000488.

Axmacher N., Cohen M.X., Fell J. et al. (2010). Intracranial EEG correlates of expectancy and memory formation in the human hippocampus and nucleus accumbens. *Neuron*, 65, 541–49.

Baldassano C., Chen J., Zadbood A. et al. (2017). Discovering event structure in continuous narrative perception and memory. *Neuron*, 95, 709–721.e5.

Banta Lavenex P., and Lavenex P. (2009). Spatial memory and the monkey hippocampus: Not all space is created equal. *Hippocampus*, 19, 8–19.

Barnett A.J., Nguyen M., Spargo J. et al. (2023). Hippocampal-cortical interactions during event boundaries support retention of complex narrative events. *Neuron*, S0896-6273(23)00766–3.

Behrens T.E.J., Muller T.H., Whittington J.C.R. et al. (2018). What is a cognitive map? Organizing knowledge for flexible behavior. *Neuron*, 100, 490–509.

Bellmund J.L.S., Gärdenfors P., Moser E.I. et al. (2018). Navigating cognition: Spatial codes for human thinking. *Science*, 362, eaat6766.

Ben-Yakov A., and Henson R.N. (2018). The hippocampal film editor: Sensitivity and specificity to event boundaries in continuous experience. *J Neurosci*, 38, 10057–68.

Billette O.V., Ziegler G., Aruci, M. et al. (2022). Novelty-related fMRI responses of precuneus and medial temporal regions in individuals at risk for alzheimer disease. *Neurology*, 99, e775–88.

Bird C.M., Capponi C., King J.A. et al. (2010). Establishing the boundaries: The hippocampal contribution to imagining scenes. *J Neurosci*, 30, 11688–95.

Bottini R., and Doeller C.F. (2020). Knowledge across reference frames: Cognitive maps and image spaces. *Trends Cogn Sci*, 24, 606–19.

Brunec I.K., Moscovitch M., and Barense, M.D. (2018). Boundaries shape cognitive representations of spaces and events. *Trends Cogn Sci*, 22, 637–50.

Buzsaki G. (2021). *The Brain from Inside Out*. Oxford, New York: Oxford University Press.

Cohen N.J., and Eichenbaum H. (1993). *Memory, Amnesia, and the Hippocampal System*. Cambridge, MA: MIT Press.

Cohn-Sheehy B.I., Delarazan, A.I., Reagh, Z.M. et al. (2021). The hippocampus constructs narrative memories across distant events. *Curr Biol CB*, 31, 4935–4945.e7.

Cohn-Sheehy B.I., and Ranganath C. (2017). Time regained: How the human brain constructs memory for time. *Curr Opin Behav Sci*, 17, 169–77.

Corkin S. (2002). What's new with the amnesic patient H.M.? *Nat Rev Neurosci*, 3, 153–60.

Davachi L. (2006). Item, context and relational episodic encoding in humans. *Curr Opin Neurobiol*, 16, 693–700.

Diana R.A., Yonelinas, A.P., and Ranganath, C. (2007). Imaging recollection and familiarity in the medial temporal lobe: A three-component model. *Trends Cogn Sci*, 11, 379–86.

Dimsdale-Zucker, H.R., and Ranganath, C. (2019). Representational similarity analyses: A practical guide for functional MRO applications. *Handbook of Behavioral Neuroscience*, 509–25.

DuBrow, S., and Davachi, L. (2013). The influence of context boundaries on memory for the sequential order of events. *J Exp Psychol Gen*, 142, 1277–86.

DuBrow, S., Rouhani, N., Niv, Y. et al. (2017). Does mental context drift or shift? *Curr Opin Behav Sci*, 17, 141–46.

Dusek, J.A., and Eichenbaum, H. (1997). The hippocampus and memory for orderly stimulus relations. *Proc Natl Acad Sci U S A*, 94, 7109–14.

Eacott, M.J., and Gaffan, E.A. (2005). The roles of perirhinal cortex, postrhinal cortex, and the fornix in memory for objects, contexts, and events in the rat. *Q J Exp Psychol B*, 58, 202–17.

Eichenbaum, H. (2017). On the integration of space, time, and memory. *Neuron*, 95, 1007–18.

Eichenbaum, H., and Cohen, N.J. (2001). *From Conditioning to Conscious Recollection: Memory Systems of the Brain*. New York: Oxford University Press.

Eichenbaum, H., and Cohen, N.J. (2014). Can we reconcile the declarative memory and spatial navigation views on hippocampal function? *Neuron*, 83, 764–70.

Eichenbaum, H., Dudchenko, P., Wood, E. et al. (1999). The hippocampus, memory, and place cells: Is it spatial memory or a memory space? *Neuron*, 23, 209–26.

Eichenbaum, H., Yonelinas, A.P., and Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annu Rev Neurosci*, 30, 123–52.

Ekstrom, A.D., and Hill, P.F. (2023). Spatial navigation and memory: A review of the similarities and differences relevant to brain models and age. *Neuron*, 111, 1037–49.

Ekstrom, A.D., and Ranganath, C. (2018). Space, time, and episodic memory: The hippocampus is all over the cognitive map. *Hippocampus*, 28, 680–87.

Epstein, R.A., Patai, E.Z., Julian, J.B., et al. (2017). The cognitive map in humans: Spatial navigation and beyond. *Nat Neurosci*, 20, 1504–13.

Ezzyat, Y., and Davachi, L. (2011). What constitutes an episode in episodic memory? *Psychol Sci*, 22, 243–52.

Farzanfar, D., Spiers, H.J., Moscovitch, M. et al. (2022). From cognitive maps to spatial schemas. *Nature Reviews Neuroscience*, 24(2), 63–79.

Franklin, N.T., Norman, K.A., Ranganath, C. et al. (2020). Structured event memory: A neuro-symbolic model of event cognition. *Psychol Rev*, 127, 327–61.

Geva-Sagiv, M., Dimsdale-Zucker, H.R., Williams, A.B. et al. (2023). Proximity to boundaries reveals spatial context representation in human hippocampal CA1. *Neuropsychologia*, 189, 108656.

Grossberg, S. (1980). How does the brain build a cognitive code. *Psychol Rev*, 87, 1–51.

Grunwald, T., Beck, H., Lehnertz, K. et al. (1999). Limbic P300s in temporal lobe epilepsy with and without Ammon's horn sclerosis. *Eur J Neurosci*, 11, 1899–906.

Hartley, T., Lever, C., Burgess, N. et al. (2014). Space in the brain: How the hippocampal formation supports spatial cognition. *Philos Trans R Soc Lond B Biol Sci*, 369. 20120510.

Hassabis, D., Kumaran, D., Summerfield, C. et al. (2017). Neuroscience-inspired artificial intelligence. *Neuron*, 95, 245–58.

Hasselmo, M.E., and McClelland, J.L. (1999). Neural Models of Memory. *Curr Opin Neurobiol*, 9, 184.

Hsieh, L.T., Gruber, M.J., Jenkins, L.J. et al. (2014). Hippocampal activity patterns carry information about objects in temporal context. *Neuron*, 81, 1165–78.

Karl, T., and Arnold, J.C. (2017). The interactive nature of cannabis and schizophrenia risk genes. In V.R. Preedy (ed.), *Handbook of Cannabis and Related Pathologies* (pp. 335–44). San Diego: Academic Press.

Kesner, R.P. (2003). Functional specificity of memory function associated with different subregions of the medial temporal lobe. *Curr Neurol Neurosci Rep*, 3, 449–51.

Kessels, R.P., de Haan, E.H., Kappelle, L.J. et al. (2001). Varieties of human spatial memory: A meta-analysis on the effects of hippocampal lesions. *Brain Res Brain Res Rev*, 35, 295–303.

Knierim, J.J., Lee, I., and Hargreaves, E.L. (2006). Hippocampal place cells: Parallel input streams, subregional processing, and implications for episodic memory. *Hippocampus*, 16, 755–64.

Kolarik B.S., Shahlaie, K., Hassan, A. et al. (2016). Impairments in precision, rather than spatial strategy, characterize performance on the virtual Morris Water Maze: A case study. *Neuropsychologia*, 80, 90–101.

Kravitz, D.J., Saleem, K.S., Baker, C.I. et al. (2011). A new neural framework for visuospatial processing. *Nat Rev Neurosci*, 12, 217–30.

Lisman, J., Grace, A.A., and Duzel, E. (2011). A neoHebbian framework for episodic memory: role of dopamine-dependent late LTP. *Trends Neurosci*, 34, 536–47.

Lisman, J.E., and Grace, A.A. (2005). The hippocampal-VTA loop: Controlling the entry of information into long-term memory. *Neuron*, 46, 703–13.

Lu, Q., Hasson, U., and Norman, K.A. (2022). A neural network model of when to retrieve and encode episodic memories. *eLife*, 11, e74445.

Manning, J., Norman, K.A., and Kahana, M.J. (2015). The role of context in episodic memory. in *The Cognitive Neurosciences, Fifth Edition*. MIT Press.

Marr, D. (1969). A theory of cerebellar cortex. *J Physiol Lond*, 202, 437–70.

Marr, D. (1970). A theory for cerebral neocortex. *Proc R Soc Lond B Biol Sci*, 176, 161–234.

Marr, D. (1971). Simple memory: A theory for archicortex. *Phil Trans R Soc Lond*, 262, 23–81.

Maurer, A.P., and Nadel, L. (2021). The continuity of context: A role for the hippocampus. *Trends Cogn Sci*. doi: 10.1016/j.tics.2020.12.007

Mayes, A., Montaldi, D., and Migo, E. (2007). Associative memory and the medial temporal lobes. *Trends Cogn Sci*, 11, 126–35.

Mayes, A.R., Meudell, P.R., and Pickering, A. (1985). Is organic amnesia caused by a selective deficit in remembering contextual information? *Cortex*, 21, 167–202.

Mayes, A.R., and Montaldi, D. (2001). Exploring the neural bases of episodic and semantic memory: The role of structural and functional neuroimaging. *Neurosci Biobehav Rev*, 25, 555–73.

McClelland, J.L., McNaughton, B.L., and O'Reilly, R.C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychol Rev*, 102, 419–57.

McCloskey, M., and Cohen, N.J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In G.H. Bower (ed.), *The Psychology of Learning and Motivation, Vol. 24* (pp. 109–64). San Diego, CA: Academic Press.

Mishkin, M., Suzuki, W., Gadian, D.G. et al. (1997). Hierarchical organization of cognitive memory. *Philos Trans R Soc Lond B*, 352, 1461–67.

Moore, J.J., Cushman, J.D., Acharya, L. et al. (2021). Linking hippocampal multiplexed tuning: Hebbian plasticity and navigation. *Nature*, 599, 442–48.

Morris, R.G. (2003). Long-term potentiation and memory. *Philos Trans R Soc Lond B Biol Sci*, 358, 643–47.

Moscovitch, M., Cabeza, R., Winocur, G. et al. (2016). Episodic memory and beyond: The hippocampus and neocortex in transformation. *Annu Rev Psychol*, 67, 105–34.

Nadel, L., and Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Curr Opin Neurobiol*, 7, 217.

Nilssen, E.S., Doan, T.P., Nigro, M.J. et al. (2019). Neurons and networks in the entorhinal cortex: A reappraisal of the lateral and medial entorhinal subdivisions mediating parallel cortical pathways. *Hippocampus*, 29, 1238–54.

O'Keefe, J., and Nadel, L. (1978). *The Hippocampus as a Cognitive Map*. London: Oxford University Press.

O'Keefe, J., and Nadel, L. (1979). Precis of O'Keefe and Nadel's *The Hippocampus as a Cognitive Map*. *Brain Behav Sci*, 2, 487–533.

O'Reilly, R.C., and McClelland, J.L. (1994). Hippocampal conjunctive encoding, storage, and recall: Avoiding a tradeoff. *Hippocampus*, 4, 661–82.

Ormond, J., and O'Keefe, J. (2022). Hippocampal place cells have goal-oriented vector fields during navigation. *Nature*, 607, 741–46.

Park, S.A., Miller, D.S., Nili, H. et al. (2020). Map making: Constructing, combining, and inferring on abstract cognitive maps. *Neuron*. doi: 10.1016/j.neuron.2020.06.030

Peer, M., and Epstein, R.A. (2021). The human brain uses spatial schemas to represent segmented environments. *Curr Biol*. doi: 10.1016/j.cub.2021.08.012

Pu, Y., Kong, X.-Z., Ranganath, C. et al. (2022). Event boundaries shape temporal organization of memory by resetting temporal context. *Nat Commun*, 13, 622.

Ranganath, C. (2010). A unified framework for the functional organization of the medial temporal lobes and the phenomenology of episodic memory. *Hippocampus*, 20, 1263–90.

Ranganath, C. (2019). Time, memory, and the legacy of Howard Eichenbaum. *Hippocampus*, 29, 146–61.

Ranganath, C. (2022). What is episodic memory and how do we use it? *Trends Cogn Sci*, 26, 1059–61.

Ranganath, C., and Ekstrom, A.D. (2020). Maps, memories, and the hippocampus. *The Cognitive Neurosciences*. MIT Press.

Ranganath, C., and Rainer, G. (2003). Neural mechanisms for detecting and remembering novel events. *Nat Rev Neurosci*, 4, 193–202.

Ranganath, C., and Ritchey, M. (2012). Two cortical systems for memory-guided behaviour. *Nat Rev Neurosci*, 13, 713–26.

Reagh, Z.M., Delarazan, A.I., Garber, A. et al. (2020). Aging alters neural activity at event boundaries in the hippocampus and Posterior Medial network. *Nat Commun*, 11, 3980.

Reagh, Z.M., and Ranganath, C. (2023). Flexible reuse of cortico-hippocampal representations during encoding and recall of naturalistic events. *Nat Commun*, 14, 1279.

Rinck, M. (2005). Spatial situation models. In A. Miyake and P. Shah (eds.), *The Cambridge Handbook of Visuospatial Thinking* (pp. 334–82). Cambridge: Cambridge University Press.

Schapiro, A.C., Turk-Browne, N.B., Botvinick, M.M. et al. (2017). Complementary learning systems within the hippocampus: A neural network modelling approach to reconciling episodic memory with statistical learning. *Phil Trans R Soc B*, 372, 20160049.

Schiller, D., Eichenbaum, H., Buffalo, E.A. et al. (2015). Memory and space: Towards an understanding of the cognitive map. *J Neurosci Off J Soc Neurosci*, 35, 13904–11.

Schlichting, M.L., and Preston, A.R. (2015). Memory integration: Neural mechanisms and implications for behavior. *Curr Opin Behav Sci*, 1, 1–8.

Scoville, W.B., and Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *J Neurol Neurosurg Psychiatry*, 20, 11–21.

Shen, K., Bezgin, G., Selvam, R. et al. (2016). An anatomical interface between memory and oculomotor systems. *J Cogn Neurosci*, 28, 1772–83.

Shohamy, D., and Wagner, A.D. (2009). Integrative encoding. *Am J Psychiatry*, 166, 284.

Smith, S.M., and Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychon Bull Rev*, 8, 203–20.

Spiers, H.J., Hayman, R.M.A., Jovalekic, A. et al. (2015). Place field repetition and purely local remapping in a multicompartment environment. *Cereb Cortex N Y NY*, 25, 10–25.

Sugar, J., and Moser, M.-B. (2019). Episodic memory: Neuronal codes for what, where, and when. *Hippocampus*, 29, 1190–205.

Teyler, T.J., and DiScenna, P. (1986). The hippocampal memory indexing theory. *Behav Neurosci*, 100, 147–54.

Teyler, T.J., and Rudy, J.W. (2007). The hippocampal indexing theory and episodic memory: Updating the index. *Hippocampus*, 17(12), 1158–1169.

Tolman, E.C. (1948). Cognitive maps in rats and men. *Psychol Rev*, 55, 189–208.

Tolman, E.C., Ritchie, B.F., and Kalish, D. (1946). Studies in spatial learning: II. Place learning versus response. *J Exp Psychol*, 36, 221–29.

Tsao, A., Sugar, J., Lu, L. et al. (2018). Integrating time from experience in the lateral entorhinal cortex. *Nature*, 561, 57–62.

Tulving, E. (1972). Episodic and semantic memory. In *Organization of Memory* (pp. xiii, 423). Oxford: Academic Press.

Umbach, G., Kantak, P., Jacobs, J. et al. (2020). Time cells in the human hippocampus and entorhinal cortex support episodic memory. *Proc Natl Acad Sci U S A*, 117, 28463–74.

Wigström, H. (1973). A neuron model with learning capability and its relation to mechanisms of association. *Kybernetik*, 12, 204–15.

Yonelinas, A.P., Ranganath, C., Ekstrom, A.D. et al. (2019). A contextual binding theory of episodic memory: Systems consolidation reconsidered. *Nat Rev Neurosci*, 20, 364–75.

Zacks, J.M. (2020). Event perception and memory. *Annu Rev Psychol*, 71, 165–91.

Zheng, J., Schjetnan, A.G.P., Yebra, M. et al. (2022). Neurons detect cognitive boundaries to structure episodic memories in humans. *Nat Neurosci*, 25, 358–68.

Zucker, H.R., and Ranganath, C. (2015). Navigating the human hippocampus without a GPS. *Hippocampus*, 25, 697–703.

8

Coding of Space and Time for Memory Function

Michael E. Hasselmo, Jennifer C. Robinson, Patrick A. LaChance, Jacob H. Wilmot, L. Kelton Wilmerding, Samantha Malmberg, Mahir Patel, Quan Do, and G. William Chapman

Introduction

Multiple cortical structures are implicated in episodic memory function. This review will present data on the neuronal activity that codes space and time in specific cortical structures, including both entorhinal cortex and hippocampus. These data have been used in models of episodic memory that involve the encoding and retrieval of spatiotemporal trajectories. This review will also present data and modelling of the transformation between egocentric and allocentric coordinate systems.

Episodic memory was defined as memory for events that occur at a specific place and a specific time (Tulving 1984; Eichenbaum et al. 1999). Episodic memory was described as 'snapshots whose orderly succession can create the mnemonic illusion of the flow of past time' (Tulving 1984). A neural model of episodic memory (Hasselmo 2009; Hasselmo 2012) proposed that episodic memory must contain more than a temporal sequence of snapshots, but instead include a continuous representation of time and space as a spatiotemporal trajectory (Figure 8.1), that includes the speed and direction of movement of an agent or other objects (Hasselmo et al. 2010; Hasselmo 2012), the agent's viewpoint of an event (Conway 2009), and the coding of prior context for disambiguation of memories (Hasselmo and Eichenbaum 2005; Hasselmo 2009). This chapter will review neural data relevant to modelling of episodic memory, as well as briefly reviewing models and emphasizing approaches that include modelling of the transformation between egocentric and allocentric coordinate systems. Finally, there will be a section emphasizing the importance of mathematical models to move beyond verbal definitions.

Coding of Time (Time Cells)

Episodic memory can be succinctly defined as 'What did you do at time T in place P' (Tulving 1984). The neural coding of time and space is important for

Michael E. Hasselmo et al., *Coding of Space and Time for Memory Function*. In: *Space, Time, and Memory*. Edited by: Lynn Nadel and Sara Aronowitz, Oxford University Press. © Oxford University Press (2025). DOI: 10.1093/oso/9780192882547.003.0008

this definition, and many neurons code both time and spatial location in the hippocampus (Pastalkova et al. 2008; MacDonald et al. 2011; Kraus et al. 2013), and the entorhinal cortex (Kraus et al. 2015; Bright et al. 2020). Neurophysiological data shows neurons, termed 'time cells', that code time intervals relative to task events, such as the onset of a delay period, usually in a task on which the animal is extensively trained. In many tasks used to study time cells, a rodent runs around a loop of elevated track, and then remains in a consistent single location and direction while running for 10–20 seconds on a running wheel (Pastalkova et al. 2008) or a treadmill (Kraus et al. 2013; Kraus et al. 2015). Many hippocampal and entorhinal cells will fire reliably as the animal runs through specific spatial locations on the elevated track. In addition, during the delay period of running in one location, individual time cells fire at specific time intervals after the onset of the delay. Running is not essential as time cells can also appear during delay periods of a delayed matching task performed without running (MacDonald et al. 2011) or even during a delay period in stationary head fixed animals (MacDonald et al. 2013; Heys and Dombeck 2018). Time cell responses have been shown in a wide range of different structures, including hippocampal region CA1 (Pastalkova et al. 2008; Kraus et al. 2013; Mau et al. 2018), hippocampal region CA3 (Salz et al. 2016) as well as the entorhinal cortex (Kraus et al. 2015; Tsao et al. 2018; Heys and Dombeck 2018). Time cell responses also appear in prefrontal cortex, but without such clear evidence of location coding (Tiganj et al. 2017). These data show that time cell responses could help in disambiguating events that occur at one time point versus another on a time scale of seconds.

More recent data shows that neurons also show a difference in calcium activity across the trials within a given day (Mau et al. 2018), and some show consistencies across days supporting coding of time on the time scale of minutes (MacDonald et al. 2011; Liu et al. 2022). This supports a model of multiscale logarithmic coding of time developed by Marc Howard and colleagues (Howard et al. 2014). Calcium imaging of the same population of hippocampal neurons over several days (Mau et al. 2018) also shows that time cells drop out or appear slowly over days, resulting in a change in correlation across the population on time scales of days (Mau et al. 2018). This slow drift in ensemble membership could provide a differential coding of memories on different days (Howard et al. 2014; Mankin et al. 2015; Rubin et al. 2015; Cai et al. 2016; Rule et al. 2019; Kinsky et al. 2020; Ziv et al. 2013; Levy et al. 2021). Models show that changes in the correlation of a population of cells on multiple longer temporal scales are essential for the capacity to differentiate episodic memories occurring at different time points on the scale of seconds, minutes, and hours (Howard et al. 2014; Liu et al. 2019). This multiscale representation provides an efficient representation for memory on different scales, but it is also possible that time coding involves different mechanisms at different scales (Phillips 2014), with time cells coding seconds, minutes, and possibly hours, but the slow drift providing coding for long scales on the order of days, weeks, months, or years. The focus on disambiguating different time points motivated the use of the term 'time cell' (Howard and Eichenbaum 2013; Eichenbaum 2014; Kraus et al. 2013;

MacDonald et al. 2011). However, these cells could be coding sequences of associations of internal and external features making up the events in an episode (Buzsaki and Tingley 2018) consistent with the initial description of these neurons as 'episode cells' (Pastalkova et al. 2008).

Combined Coding of Time, Distance, and Location

One experiment compared time versus running distance by recording as a rat ran on a treadmill at different speeds during different delay periods (Kraus et al. 2013). This experiment showed that neurons could respond on the basis of either time or running distance during the delay (Kraus et al. 2013) as predicted by previous models (Hasselmo 2008; Burgess et al. 2007). Neurons that fire as time cells also code other dimensions related to episodic memory, consistent with evidence of mixed selectivity in other regions (Rigotti et al. 2013). Some cells that fire as time cells during running on the treadmill also fire as place cells during running off the treadmill on the return arms (Kraus et al. 2013; Mau et al. 2018) indicating that these cells do not only code time but combine location and time in an episodic trajectory. Coding of both space and time was also shown for single grid cells in the entorhinal cortex (Kraus et al. 2015). These grid cells fired in an array of spatial locations when animals foraged in a two-dimensional environment, but also fired as time cells at different time points during a 16 second delay as the rat ran in a single location on the treadmill. Interestingly, a recent study showed that the overall population of space-encoding neurons and neurons coding time during immobility in entorhinal cortex form anatomically distinct sub-populations (Heys and Dombeck 2018; Heys et al. 2014). The coding of episodes as spatiotemporal trajectories only requires a population code containing both space and time, so this model can function whether the representation is shared across individual neurons or appears in different populations (Hasselmo 2012). Human imaging data shows neural activity in hippocampus and parahippocampal regions associated with disambiguating the retrieval of overlapping trajectories in virtual mazes (Brown and Stern 2014; Brown et al. 2010; Brown et al. 2014).

Phase Coding of Time

Time cells code time not only by overall firing rate, but also by the phase of firing of cells relative to rhythmic oscillations at theta frequency in the local field potential. Time cells show theta phase precession within their firing fields (Pastalkova et al. 2008; Terada et al. 2017; Ning et al. 2022). A a time cell starts firing, its spikes appear at late phases of theta, and as time evolves the spiking shifts to earlier phases of theta cycle before spiking ends. The coding of space by place cells shows a similar shift in phase as an animal runs through a place field. This phase coding appears important for the temporal specificity of time cell firing as time cells lose their

time selectivity during inactivation of the medial septum (Wang et al. 2015), which reduces theta rhythm in the hippocampus (Brandon et al. 2014; Rawlins et al. 1979). The same inactivation of medial septum also removes spatial specificity of grid cell firing (Brandon et al. 2011; Koenig et al. 2011). Theta phase coding has the advantage that it could allow a single neuron to code a continuous dimension of time or space, which might allow a broader range of transformations on the level of single neurons that might be difficult to implement across a full population.

Coding of Spatial Location

Place Cells in Hippocampus

Neurophysiological recording in the hippocampus revealed place cells that fire based on the spatial location of the animal being recorded (O'Keefe 1976; O'Keefe and Dostrovsky 1971). The potential role of these place cell responses for behavioural function, and their relationship to philosophical questions about the a priori representation of space were addressed extensively in the influential book by O'Keefe and Nadel on coding of space as a cognitive map (O'Keefe and Nadel 1978). The new perspective provided by these experimental findings and the cognitive map theory could be seen as causing a paradigm shift within neuroscience from a focus on operant conditioning experiments to a focus on the more continuous dimensions of spatial behaviour. This work engendered numerous subsequent studies that showed place cell responses when animals were in specific locations during foraging an open field environments (Muller et al. 1987; O'Keefe and Burgess 1996; Huxter et al. 2008; Lever et al. 2002), and in local areas of linear tracks (O'Keefe and Recce 1993), the 8-arm radial maze (McNaughton et al. 1983), or a spatial alternation task (Ainge et al. 2007; Wood et al. 2000; Kinsky et al. 2020). The firing of place cells can vary dependent on many factors including direction through the place field (Fenton and Muller 1998; Redish 1999) and cells can have more than one firing field (Fenton et al. 2008). The position of an animal can be effectively decoded from the firing activity of hippocampal place cells (Brown et al. 1998), supporting their role in guiding behaviour in spatial memory tasks. In more recent studies, hippocampal cell firing has been shown to code information about the direction and distance of goal locations (Ormond and O'Keefe 2022).

Phase Coding by Place Cells

In addition to coding location by firing rate, place cells also code location by their phase of firing in a phenomenon called theta phase precession (O'Keefe and Recce 1993). This discovery motivated testing for phase precession when time cells were discovered much later. As the animal enters the firing field of a place cell, spiking occurs at late phases of theta and then shifts to earlier phases as the animal

runs through the firing field and exits (O'Keefe and Recce 1993; Skaggs et al. 1996; Maurer et al. 2006; Schmidt et al. 2009; Zugaro et al. 2005). Theta phase precession is associated with sequential spiking of neurons coding sequential places on different phases (Foster and Wilson, 2007), but theta phase precession appears on the first trial of running on a novel linear track, whereas theta sequences only appear on later trials (Feng et al. 2015). These data suggest a role of spike timing and phase for the episodic coding of spatiotemporal trajectories.

Grid Cells in Entorhinal Cortex

Neurophysiological recording in the entorhinal cortex demonstrates different types of coding of spatial dimensions that differ from place cells. One striking form of coding involves the activity of entorhinal grid cells, which spike in a hexagonal array of spatial locations as an animal forages broadly (Hafting et al. 2005). Different grid cells fire with different size and spacing between firing fields, allowing a population of grid cells to code a single location (Sargolini et al. 2006; Barry et al. 2007; Stensola et al. 2012; Heys et al. 2014). Many grid cells code both for the animals location and the current head direction of the animal (Sargolini et al. 2006), consistent with the joint coding of location and direction in a spatiotemporal trajectory (Hasselmo 2012). In addition other neurons code head direction alone, which is not found in the hippocampus, but had been shown previously in structures providing input to entorhinal cortex including the postsubiculum (Taube et al. 1990) and the anterior thalamus (Taube 1995).

Phase Coding by Grid Cells

Grid cells also exhibit phase coding in the form of theta phase precession by grid cells in layer II of entorhinal cortex as an animal runs on a linear track (Hafting et al. 2008), or as an animal forages in two dimensions in an open field (Climer et al. 2013; Jeewajee et al. 2014). Consistent with this, the intrinsic rhythmicity of entorhinal neurons differs with spatial scale (Jeewajee et al. 2008) and shifts with running speed (Hinman et al. 2016; Jeewajee et al. 2008). In contrast to layer II neurons, layer III grid cells show phase locking to theta rhythm. The potential role of theta rhythm in generation of grid cell responses is supported by the fact that inactivation of the medial septum causes both a dramatic reduction of theta rhythm in the entorhinal cortex (Mitchell et al. 1982; Jeffery et al. 1995), and a loss of the spatial selectivity of firing of grid cells (Brandon et al. 2011; Koenig et al. 2011). Inactivation of the medial septum or lesions of the fornix also cause impairments in the memory of goal locations (O'Keefe et al. 1975; Chrobak et al. 1989; Brioni et al. 1990). The specific population of medial septal neurons involved in regulating grid cell firing has not yet been demonstrated. However, recent studies show that inactivation of glutamatergic neurons causes a decreased specificity in grid cell firing activity, and inactivation of

GABAergic neurons in the medial septum results in the loss of grid cell spatial firing along with a reduction in theta rhythm oscillations (Robinson et al. 2019).

Behavioural Function of Place Cells and Grid Cells

Studies support a role for place cells and grid cells in spatial navigation. A range of studies show that damage to the hippocampus causes impairments in two-dimensional spatial navigation tasks such as the Morris water maze, in which the animal learns a specific goal platform location and must then navigate to that location from a range of different starting locations (Morris et al. 1982; Eichenbaum et al. 1990). Models have addressed how different neural subtypes could underlie planning of spatiotemporal trajectories to generate the correct trajectory to the goal location from a new start location (Erdem and Hasselmo 2014; Erdem and Hasselmo 2012; Redish and Touretzky 1998). Early models focused on the role of place cells and head direction cells (Redish and Touretzky 1998), but those models required dense representation of spatial locations by place cells. Later models address the additional role of grid cells and speed cells in planning trajectories to a goal location without needing to form a place cell code for each location (Erdem and Hasselmo 2014; Erdem and Hasselmo 2012; Kubie and Fenton 2012). Recent models have addressed how spatial coding could reflect a successor representation of possible future states for effective reinforcement learning (Momennejad 2020; Brunec and Momennejad 2022).

Impairments of goal-finding in the Morris water maze are also observed after lesions of the entorhinal cortex (Steffenach et al. 2005) or the dorsal presubiculum (Taube et al. 1992), consistent with a role of different neuronal subtypes such as head direction cells and grid cells from these regions. Further data shows neural coding of position along a trajectory in structures such as the retrosplenial cortex (Alexander and Nitz 2017; Alexander and Nitz 2015), and human imaging data shows coding of arc length along a trajectory in addition to coding of euclidean distance, translation and rotation (Chrastil et al. 2016; Chrastil et al. 2015).

Coding of Prior Context

The performance of tasks such as delayed spatial alternation or delayed non-match to position requires the capacity to distinguish (disambiguate) spatial location on different trials, as described in previous modelling work (Levy 1996; Hasselmo and Eichenbaum 2005; Hasselmo 2009). Lesions of the hippocampus cause impairments in these types of tasks (Ainge et al. 2007; Aggleton et al. 1986; Aggleton et al. 1995; Hallock et al. 2013; Emerich and Walsh 1989; Costa et al. 2005) that require memory of both spatial location and the specific time of the prior trial to disambiguate it from other previous trials. Neurophysiological data in related continuous

alternation tasks (without a delay) show context-dependent activity appropriate for this behavioural disambiguation based on memory. For example, when a rat runs on the stem of continuous spatial alternation task, individual neurons will fire selectively based on the past or future turning response. These 'splitter' neurons have been observed in the hippocampus (Wood et al. 2000; Ferbinteanu and Shapiro 2003; Kinsky et al. 2020; Levy et al. 2021) and entorhinal cortex (Frank et al. 2000; Lipton et al. 2007; O'Neill et al. 2017). The separation and disambiguation of overlapping spatiotemporal trajectories is a challenge for any model of episodic memory, particularly when considering the large number of memories that can be generated in a single familiar environment (Robins 2015). This raises questions of the relative overlap of representations, which could range from the extreme case of an index model, in which a small number of non-overlapping neurons code each memory, to the alternate case of a broadly distributed representation in which most neurons are involved in every memory, which is more amenable to cueing of memories, but less amenable to prevention of interference (Robins 2015). The nature of the amount of overlap and capacity to cue less overlapping memories remains an empirical question for both simulations and neurophysiology (Hasselmo 2015).

The context-dependent neuronal responses in a spatial alternation task can appear at specific times during training and are more stable than place cell responses in the task, possibly because the splitter responses are more necessary for accurate task performance (Kinsky et al. 2020). The left-right discriminability of splitter cell responses correlates significantly with accurate behavioural performance (Kinsky et al. 2020). Context-dependent activity can also distinguish the sample versus test trials in delayed non-match to position (Griffin et al. 2007; Levy et al. 2021), and during the course of learning the task shows a gradual shift from coding both turn direction and task phase, to showing more coding of turn direction or task phase alone (Levy et al. 2021). The separation of representation could also include the dentate gyrus, as neural activity associated with turning to one side of the maze differs from the representation associated with turning both directions (Wilmerding et al. 2023). These studies are consistent with impairments of delayed-non-match-to-position (DNMP) caused by dentate gyrus lesions (Emerich and Walsh 1989; Costa et al. 2005), and with earlier studies showing splitting of different spatial maps over time (Redish 1999).

Both the guidance of behaviour and the learning-dependent shift in context-dependent representations over time may depend upon mechanisms of sequence retrieval during theta rhythm. Theta sequences appear to reflect planning of future trajectories, as sequences appear at choice points (Johnson and Redish 2007; Kay et al. 2020), and the length of theta sequences increases with greater distance of future goals (Wikenheiser and Redish, 2015). The phase of firing relative to theta rhythm also appears to shift based on the novelty of individual cues (Manns et al. 2007) or the novelty of the environment (Wells et al. 2013; Douchamps et al. 2013), consistent with proposals for encoding and retrieval on different phases of theta rhythm cycles (Hasselmo et al. 2002; Hasselmo 2006). If encoding and retrieval

processes occur on different phases of theta, then the ability to discriminate a retrieved memory from current sensory input could depend upon an intact theta rhythm (Hasselmo 2005). Loss of this phase coding could result in confabulation of imaginary memory with real memory, which has been shown to occur after damage to the medial septum (DeLuca and Cicerone 1991). The different phases of encoding and retrieval could include a neural signal associated with phase that distinguishes the memory of a real event from the memory of an imagined event (Boyle 2021).

Coding of Trajectory Speed and Direction

In contrast to the model of episodic memory as a series of snapshots, the model of episodic memory as a continuous spatiotemporal trajectory (Figure 8.1) includes dimensions beyond spatial location and time interval (Hasselmo 2012; Hasselmo 2009). The fact that one can remember a specific viewpoint of a scene (Conway 2009), or changes in speed of movement indicate that direction and speed are available in episodic memories.

In addition to the importance of speed and direction for episodic memory, many models of grid cell firing use path integration of self-motion to code location. This standard mechanism could function in parallel with coding of location by transformation of sensory input. Grid cells retain their spatial firing pattern in darkness, suggesting a role for path integration of self-motion in the absence of visual cues (Hafting et al. 2005; Dannenberg et al. 2020). However, sensory input is important as grid cells rotate with visual cues in a circular environment (Hafting et al. 2005), and spatial coding by grid cells is lost when all sensory cues are lost or obscured including visual, auditory, somatosensory, and olfactory input (Chen et al. 2016; Pérez-Escobar et al. 2016). The role of sensory input is further supported by evidence that grid cells lose spatial coding during inactivation of regions providing head direction input (Winter et al. 2015). The subsequent sections will review data on the potential role of memory for spatial location based on path integration versus the transformation of sensory input.

Coding of Speed

Path integration involves integration of velocity, which would consist of movement speed and direction. The running speed of animals has been shown to be coded by neurons in the hippocampus (O'Keefe et al. 1998; McNaughton et al. 1983) and medial entorhinal cortex (Sargolini et al. 2006; Wills et al. 2012; Buetfering et al. 2014; Kropff et al. 2015; Hinman et al. 2016). Some cells appear to selectively code running speed (Kropff et al. 2015), but others show mixed selectivity as grid cells and head direction cells that also code running speed (Sargolini et al. 2006; Wills et al. 2012; Buetfering et al. 2014; Jeewajee et al. 2008; Hinman et al. 2016). Coding of running speed also appears in retrosplenial and parietal cortex (McNaughton

et al. 1994; Alexander et al. 2020; Clancy et al. 2019; Carstensen et al. 2021) and sensory responses are modulated by running speed in visual cortex (Niell and Stryker 2010) and auditory cortex (Nelson and Mooney 2016). The direct coding of speed could occur without requiring the computation of distance over time, as neurons also show responses to acceleration guided by the vestibular system, such that integration of acceleration could generate a speed response independent of the ongoing perception of distance or time.

This speed coding could be important for episodic memory of spatiotemporal trajectories. In contrast, the data on speed coding does not uniformly support the use of speed for path integration of two-dimensional spatial location. Models of path integration require linear coding of running speed by firing rate, as found in many speed tuning curves, but many speed-modulated cells show non-linear responses that saturate at moderate speeds (Hinman et al. 2016; Dannenberg et al. 2019). Surprisingly, neurons represent speed by firing rate over intervals of several seconds, but over shorter periods than a second, the firing rate code is too inaccurate for effective path integration (Dannenberg et al. 2019). This calls into question the use of coding of location based on integration of a firing rate code for running speed.

Coding of Direction

In addition to coding of running speed, both models of episodic memory as spatiotemporal trajectories and path integration models of grid cells require coding of movement direction. The data on coding of movement direction is less extensive and less consistent with path integration than the data on running speed. Many place cells show sensitivity to movement direction on a one-dimensional linear track (McNaughton et al. 1983; Huxter et al. 2003), though this may reflect the change in goal points for different directions of running (Redish 1999; Jackson and Redish 2007). Hippocampal cells also appear to code the direction of a goal on a honeycomb maze (Ormond and O'Keefe 2022). Outside the hippocampus, there are many neurons that show responses to the current allocentric direction of an animal's head (Taube et al. 1990). These head direction cells do not depend on current location or movement direction. A systematic analysis of neurons in entorhinal cortex during periods when movement direction differed from head direction demonstrated numerous head direction cells but no cells that exclusively code movement direction (Raudies et al. 2015). Researchers have proposed that movement direction could instead be coded by rhythmic firing of theta cells (Welday et al. 2011) or spiking in theta sequences (Zutshi et al. 2017). Theta phase precession codes movement direction with a shift from late to early phases as animals move backwards (Maurer et al. 2014) or ride backwards on a train (Cei et al. 2014). However, the broad evidence for head direction cells in the absence of movement direction cells suggest that path integration of self-motion might be less important than the coding of sensory feature angle provided by head direction cells (Raudies et al. 2015).

Head direction cells could instead be vital for accurate transformation of egocentric coordinates into allocentric coordinates, as allocentric head direction can be used to transform egocentric coordinates of sensory feature angle into allocentric location (Byrne et al. 2007; Touretzky and Redish 1996; Bicanski and Burgess 2018). Head direction cells have been found in a range of structures including dorsal presubiculum (Taube et al. 1990), anterior thalamus (Taube 1995), and entorhinal cortex (Sargolini et al. 2006; Brandon et al. 2013; Brandon et al. 2011; Giocomo et al. 2014). Lesions of the dorsal presubiculum or anterior thalamic nucleus, which both provide head direction input to cortex, cause destablization of hippocampal place cells (Goodridge and Taube 1997) and loss of spatial coding by entorhinal grid cells (Winter et al. 2015). Head direction cells usually do not show theta rhythmicity, but in the entorhinal cortex these cells can show theta rhythmic firing that falls on alternate cycles of the theta rhythm (Brandon et al. 2013), consistent with place cell readout of trajectories on alternate theta cycles in the hippocampus (Kay et al. 2020).

Coding Based on Different Coordinate Systems

The above data indicates that path integration of self-motion may not be the most important mechanisms for updating the memory of spatial location. In contrast, the influence of sensory feature angle could be used to update the memory of spatial location, as supported by the influence of visual cue rotation on the firing of place cells (Muller and Kubie 1987) when cues are stable (Knierim et al. 1998), and on grid cells (Hafting et al. 2005) and the loss of grid cell firing when darkness is combined with removal of auditory and somatosensory cues (Chen et al. 2016; Pérez-Escobar et al. 2016). Understanding the influence of sensory input on place cell firing requires understanding of coordinate transformations. The coding of spatial location by place cells is commonly described in allocentric coordinates (i.e., allocentric coordinates describe the position of an animal relative to environment boundaries). This requires a coordinate transformation from the egocentric coordinates of sensory input such as visual feature angle (i.e., egocentric coordinates describe the position of a feature relative to an animal). This may correspond to the philosophical description of a perspectival view of an object compared to a constant representation of an object (Green and Schellenberg 2017). Studies of neural activity in the human brain have demonstrated differences in neural activity associated with viewing a navigation task from a first-person, egocentric perspective, compared to performing the task from a third-person overhead perspective (Sherrill et al. 2013; Sherrill et al. 2015). The following sections will review further data relevant to this topic.

Allocentric Boundary Cells

The important influence of sensory cues for location coding was shown in experiments in which the distance of environmental barriers relative to other barriers (i.e., allocentric position) was changed (i.e., changing a 1×1 metre square environment to a 1×2 metre rectangle) and this was shown to alter the position of the firing fields of place cells (O'Keefe and Burgess 1996). This data motivated the theoretical proposal of boundary vector cells that code animal position relative to boundaries (Burgess et al. 2000; Hartley et al. 2014; Hartley et al. 2000). This theoretical prediction was later supported by data demonstrating boundary vector cells that fire when boundaries are at a specific distance and allocentric angle in allocentric coordinates (Solstad et al. 2008; Lever et al. 2009; Savelli et al. 2008; Barry et al. 2006; Poulter et al. 2021). These boundary responses can occur at a distance from the boundary (Lever et al. 2009), indicating a role of visual sensory cues, and they show responses to insertion of new barriers in the environment (Lever et al. 2009; Poulter et al. 2021). The responses can also persist after removal of an inserted barrier (Poulter et al. 2021). A related population of neurons, termed object vector cells, spike when the animal occupies specific allocentric angles and distances from non-boundary objects (Hoydal et al. 2019; Deshmukh and Knierim 2011; Deshmukh and Knierim 2013). Grid cells also alter activity based on the allocentric positions of environmental boundaries, showing compression or expansion of the distance between firing fields (Barry et al. 2007; Stensola et al. 2012; Munn et al. 2020), and changes in coding of velocity with wall movement (Munn et al. 2020).

Models have demonstrated how allocentric spatial location can be generated from egocentric visual coding of boundaries (Byrne et al. 2007; Bicanski and Burgess 2018; LaChance and Taube 2023; O'Keefe 1990). Models have demonstrated how allocentric boundary vector cells could be generated from egocentric sensory coding of environment boundaries combined with head direction (Burgess et al. 2000; Hartley et al. 2000; Byrne et al. 2007; Bicanski and Burgess 2018). These latter models predicted the existence of neurons that code the egocentric position of boundaries. Influences of boundary location on grid cells and allocentric boundary cells has also been modelled based on the angle and optic flow of visual features (Raudies and Hasselmo 2015; Sherrill et al. 2015) or based on matching of egocentric sensory input with a stored representation of input (Alexander et al. 2023).

Egocentric Boundary Cells

The prediction that allocentric boundary cells could be generated from egocentric boundary cells was supported by data showing egocentric boundary cells in

a range of structures, including the retrosplenial cortex (Alexander et al. 2020; van Wijngaarden et al. 2020), the postrhinal cortex (LaChance et al. 2019; Gofman et al. 2019), the entorhinal cortex (Wang et al. 2018; Wang et al. 2020) and structures receiving output from these regions such as the dorsomedial striatum (Hinman et al. 2019) and parietal cortex (Alexander et al. 2020). As animals forage in an open field environment, egocentric boundary cells fire selectively when the barriers or boundaries of the environment are at a specific angle and distance to the animal (Hinman et al. 2019; Alexander et al. 2020). They are therefore most efficiently described by plotting the position of the barrier for each spike in egocentric polar coordinates. The sum over all spikes shows how neurons respond to a barrier at a specific distance of a few to many tens of centimetres. Different neurons respond selectively for barriers at a specific range of angles and distances relative to the animal itself, with most right hemisphere neurons responding to barriers directly to the left of the animal, and most left hemisphere neurons responding to barriers directly to the right (Alexander et al. 2020). Other neurons respond to angles to the front or behind the animal. Many egocentric boundary vector responses are invariant to the appearance of environmental boundaries indicating that the response property is not driven by high-level visual features (Hinman et al. 2019). While some papers focus on the egocentric response to barriers, others focus on the coding of position relative to the centre of the environment or specific objects or goals (Wang et al. 2018; Wang et al. 2020; LaChance et al. 2019; LaChance and Taube 2023). The neurons show tuning to barriers at a number of distances, including distances well outside the range of whisker contact, as well as at a number of angles, including positions behind the animal.

Egocentric coding of the environments has also been shown in several other cortical regions, including posterior parietal cortex, secondary motor cortex, and postrhinal cortex (LaChance et al. 2019; Gofman et al. 2019; Alexander et al. 2020). In the postrhinal cortex egocentric boundary responses persist in darkness (LaChance et al. 2019), supporting the computational theory of their generation by some mechanism of path integration based on prior contact with the barrier. In the postrhinal cortex, egocentric bearing was found to be anchored to the centre of the environment rather than the boundaries (LaChance et al. 2019).

The theta phase coding shown for place cells and grid cells reviewed above suggests that theta phase coding might occur for egocentric and allocentric boundary cells. Retrosplenial neurons show phasic firing relative to hippocampal theta rhythmicity (Alexander et al. 2018) and some egocentric boundary cells show theta phase locking (Alexander et al. 2020). Consistent with the proposed separation of encoding and retrieval on different theta phases in the hippocampus (Hasselmo et al. 2002), allocentric boundary cells fire on different phases of theta during direct experience of boundaries versus trace responses to boundaries that are no longer present (Poulter et al. 2021). Theta phase coding could provide a component of spatial representation that can contribute to the transformation from egocentric

coordinates to allocentric coordinates and the encoding of egocentric information in spatiotemporal trajectories for episodic memory.

Brief Review of Hippocampal Models

The above sections reviewed some of the data on neural representations for time and space that are relevant to episodic memory. This section will provide a brief review of some existing models of episodic memory and the internal representations used in those models. Historically, both psychological and neural models of episodic memory have focused on the representation of the world as a set of vectors, usually with an arbitrary mapping of environmental features (such as words in a verbal memory task) to individual vectors. This mapping was used in many memory models in mathematical psychology (Murdock 2005), and the vector representation carried over into neural network models of episodic memory (McNaughton and Morris 1987) that focused on encoding of an array of vectors representing individual memories (McNaughton and Morris 1987). These memories were proposed to undergo orthogonalization (later called pattern separation) in the dentate gyrus (McNaughton and Morris 1987; Hasselmo and Wyble 1997; Treves and Rolls 1994), and then to be stored in a recurrent auto-associative memory in region CA3 that could mediate pattern completion (McNaughton and Morris 1987; Hasselmo and Wyble 1997; Treves and Rolls 1994). Finally, region CA1 would map the stored patterns back to the input.

One criticism of these early models is their relatively small capacity relative to the number of neurons, and also the lack of a representational structure to match space and time in the world. Another concerns the simple nature of the arbitrary vector representation, which does not reflect any hierarchical representation of information, though that could be feasible in vector representations such as those generated by convolution (Eliasmith et al. 2012; Plate 1995). These models also lack many of the elements of neural dynamics, such as network and cellular oscillations and dendritic subthreshold dynamics.

Other neural models of memory focus on the role of neurons with specific functional properties, such as place cells, grid cells, and head direction cells. These models have addressed the potential functional role of these subtypes in representing a spatial environment. Models of this type have focused on a number of functions, including: (1) guidance of behaviour to a specific goal location from a variety of start locations using place cells (Redish and Touretzky 1998; Burgess et al. 1997; Arleo et al. 2004), or using grid cells (Erdem et al. 2015; Erdem and Hasselmo 2014; Erdem and Hasselmo 2012; Kubie and Fenton 2012); (2) the encoding and retrieval of previously encoded trajectories in episodic memory (Hasselmo and Eichenbaum 2005; Hasselmo 2012; Hasselmo 2009); or (3) the transformation between egocentric and allocentric representations of environmental features

(Byrne et al. 2007; Bicanski and Burgess 2018; Alexander et al. 2023; Sheynikhovich et al. 2009). Merging these different functions in a single model would be desirable. Models that contain functional subtypes will automatically resemble those properties of the biological data, but their function for practical behaviour may be limited. In particular, many of these models require input that has already been coded in terms of specific aspects of spatial location or velocity, though a few have been designed to respond on the basis of direct egocentric sensory input (Sheynikhovich et al. 2009; Arleo et al. 2004).

The previously mentioned model of episodic memory as a spatiotemporal trajectory (Hasselmo 2012; Hasselmo 2009) effectively encodes and retrieves continuous trajectories through an environment and associated features of events at different positions along the trajectory, as summarized in Figure 8.1. During encoding, this model starts with an initial pattern of grid cell population activity that corresponds to the spatial phases of grid cells with multiple different spatial scalings (Figure 8.2). This initial pattern of grid cell activity drives firing of a population of place cells based on random excitatory connections and selection of sparse place cells with sparse firing. These place cells can be associated with the an initial egocentric view of the environment via Hebbian synaptic modification (Figure 8.2), as well as the current movement direction and running speed of the animal. As the animal continues to move, the behavioural input drives neurons coding running speed and movement direction that in turn update a model of grid cells to change their activity pattern, and this correspondingly drives a different set of place cells that become associated with the current running speed and movement direction that links to the next location code by grid cells (Figure 8.2). During retrieval, an egocentric viewpoint can drive the associated place cells that then drive the speed and direction cells to update the grid cell code in the direction of the trajectory that then updates a new set of place cells that activate the associated speed and direction cells for the next segment of the trajectory as well as associated egocentric viewpoints (Figure 8.2). In this manner, the model can retrieve extended trajectories from episodic memory (Hasselmo 2012; Hasselmo 2009). However, this model must be modified to account for data that are inconsistent with the fact that this model used grid cell input to generate place cells and also used speed-modulated direction cells that depend on movement direction rather than head direction (Hasselmo 2012; Hasselmo 2009). More recent models have suggested that a similar configuration for retrieving trajectories might have a high capacity for associative memory function (Sharma et al. 2022).

The models that start with functional cell types such as place cells and grid cells often do not account for detailed biophysical dynamics of individual neuronal conductances. More biophysically detailed models simulate the membrane conductances and single cell membrane potential dynamics of neurons (Traub et al. 2022; Kelley et al. 2021; Wallenstein and Hasselmo 1997; Traub et al. 1992; Sutton and Ascoli 2021). However, the computational demands of these biophysical models make it very difficult to simulate a wide range of functional cell types or to simulate the guidance of complex behaviours.

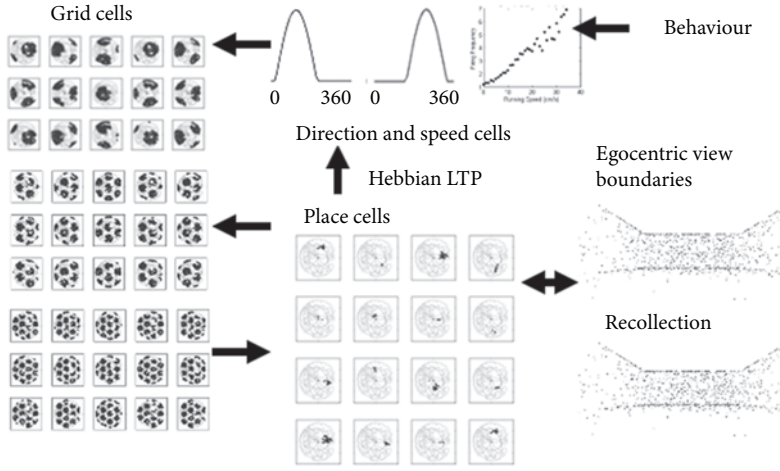


Figure 8.1 Encoding of an episodic memory as a spatiotemporal trajectory in a model (Hasselmo 2009, 2012). LEFT: A population of individual grid cells on respond based with different spatial phases and scales. MIDDLE: The grid cell population drives a set of place cells that are selected for responding a localized spatial area. TOP: The place cells form Hebbian modifications with neurons coding movement direction and running speed, that are updated by behavioural input. RIGHT: The place cells are also associated with specific egocentric viewpoints along the trajectory. Adapted with permission from Hasselmo (2012). LEFT: During encoding, the grid and place cell code form Hebbian associations with specific egocentric viewpoints representing events in the episode (different symbol shapes), while the continuous trajectory is encoded by Hebbian associations of the place code with the speed and movement direction at each point. RIGHT: During retrieval the place cells activated by an egocentric input can drive the associated direction and speed cells that update the grid cell code to drive retrieval of the trajectory and activation of place cell codes associated with other egocentric viewpoints of events along the same spatiotemporal trajectory. Adapted with permission from Hasselmo (2012)

Another approach involves using successor representations of possible future states for guiding spatial navigation using reinforcement learning (Dayan 1993). In these models, the place and grid cells arise as a model of diffusion processes for mapping potential future states (Stachenfeld et al. 2017). These successor representations can be used to more effectively plan new trajectories and are consistent with some aspects of neural data during spatial behaviour (Momennejad 2020; Brunec and Momennejad 2022). However, these models do not attempt to include representations of the cellular physiological properties of neurons.

A different class of models uses large multilayer systems of neuron-like elements with an implementation of an error-correcting learning rule such backpropagation of error in deep learning (Banino et al. 2018) or contrastive Hebbian learning (O'Reilly and Munakata 2000; Naud and Sprekeler 2018). These models have the advantage of not using an a priori representations of functional cell types, but instead generating these functional cell types within the model (Banino et al. 2018).

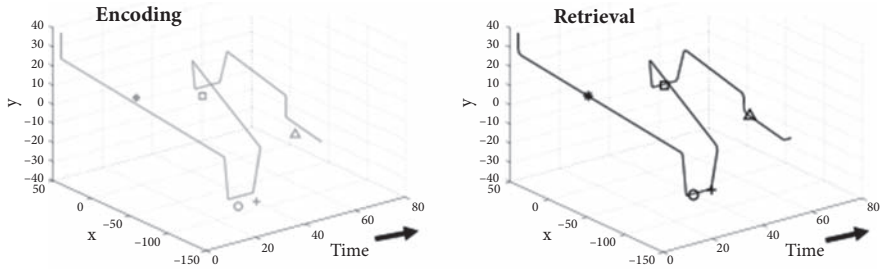


Figure 8.2 LEFT: During encoding, the grid and place cell code form Hebbian associations with specific egocentric viewpoints representing events in the episode (different symbol shapes), while the continuous trajectory is encoded by Hebbian associations of the place code with the speed and movement direction at each point. RIGHT: During retrieval the place cells activated by an egocentric input can drive the associated direction and speed cells that update the grid cell code to drive retrieval of the trajectory and activation of place cell codes associated with other egocentric viewpoints of events along the same spatiotemporal trajectory. Adapted with permission from Hasselmo (2012).

These models also have the advantage of using a more realistic egocentric input of visual input from the environment, rather than a pre-coded neural representation. Because these models generate behavioural output that usually exceeds the capabilities of most hand-wired models, this approach is considered to have great potential for advancing the understanding of neural representations. However, so far the internal dynamics of these models have not been easily interpretable to yield guidance concerning the internal dynamics that allow for the function of the models. The distributed code is difficult to decipher and does not seem to exhibit features of symbolic representations such as compositionality and productivity, that would allow breaking down their functional processes into interpretable rules and elements (Do and Hasselmo 2021). This uninterpretability also inhibits their mapping to physiological and anatomical features of real neural circuits.

In summary, the models that best simulate behaviour using error-correcting learning rules do not provide interpretable internal dynamics, but the models that directly simulate a range of functional cell types are more limited in their behavioural function and still do not address the complex dynamics of membrane potentials and membrane currents, whereas the models containing the biophysical detail of membrane currents have not been used to guide behaviour or to generate simulations of most functional cell types.

Need to Explore Broader Variety of Models

There is a strong need for better models of neural circuits. Given the limited ability of existing models to demonstrate the computational relevance of many aspects of biophysical data, one can consider the question of what would a more

effective model look like? What modifications could result in a revolution in neural memory models? The dimensions of possible models are actually enormous, and existing models have only explored a tiny fraction of this space (Hasselmo et al. 2021). Some ideas are presented here for exploring different areas of model space.

1. Phase coding. The use of temporal coding by the phase of spiking activity could provide advances over the dominant use of firing rate as the neuronal code. Most of the models described above use vectors with a continuous change in value that represents a continuous change in firing rate. These models do not directly account for the temporal coding of space and time that appears in multiple studies as theta phase precession (O'Keefe and Recce 1993; Skaggs et al. 1996; Hafting et al. 2008; Climer et al. 2013), nor do they account for the rhythmic coding of running speed (Hinman et al. 2016; Dannenberg et al. 2020). Place cells and grid cells have been modelled with phase coding based on oscillatory interference (Burgess 2008; O'Keefe and Recce 1993). However, there are difficulties for the original version of that model. Data suggests the phase code is not sensitive to disruption of internal theta (Zugaro et al. 2005), and can flow forward towards future locations (Johnson and Redish 2007; Kay et al. 2020), suggesting a stronger role in prediction of future trajectory than in path integration (Lisman and Redish 2009).
2. Dendritic processing. Another underexplored area concerns the dynamics of dendritic processing, including the potential role of phase shifts within dendritic compartments (Vaidya and Johnston 2013; Kelley et al. 2021), which could be modulated by synaptic activation of metabotropic receptors. Previous models of dendritic H current have focused on how these currents normalize timing of synaptic potential peaks (Magee 1999; Vaidya and Johnston 2013), but these currents could instead play a role in generating heterogeneity of phase based on interaction with other phase shifting inputs and matching to other dendritic inputs (Alexander et al. 2023). For example, one set of phase shifts could represent the relation between two features within an object (angle and distance) and an external phase shift could represent current object angle and current viewing angle. Differential inputs to dendritic segments has been utilized in models of phase precession (Magee 2001; Chance 2012).
3. Dendritic coding of transformations. Dendritic processing via phase shifts could allow transformations to be coded at a single neuron level rather than a network level. Many models of coordinate transformations use gain modulation (Bicanski and Burgess 2018), in which different populations of neurons code different transformations and a gating input (such as head direction) modulates the selection of different populations, requiring large numbers of neurons. In contrast, coding of transformations by individual neurons could allow coding of a broader range of transformations, and more flexible modulation of individual neuron transformations (Alexander et al. 2023).

The matching of phase codes for external egocentric input with phase codes for stored egocentric memories could allow grid cell responses based on the allocentric affine transformation (translation and rotation) between viewpoints (Alexander et al. 2023). The dynamics for phase shifting providing dendritic transformations could be stored with membrane current changes induced during plateau potentials (Grienberger et al. 2017; Bittner et al. 2017; Bittner et al. 2015).

4. Transmission of memory representations. Another appealing feature would be to have memory representations that can be easily transmitted between neurons, so that instead of requiring a long-term synaptic modification they could instead be rapidly transmitted. One possible way of doing this would be to have neural phase codes that can be transmitted between different neurons to activate cellular mechanisms that regulate phase within individual neurons. For example, neural activity could phosphorylate the H current or potassium currents in such a manner to change their time constant (Chen et al. 2001). The H current has been shown to shift the frequency and the phase of neural activity in dendrites (Vaidya and Johnston 2013; Kelley et al. 2021). Changes in these currents could underlie alterations in neuronal responses such as those contributing to place cell firing (Bittner et al. 2015).
5. Analogy with computer animation. Beyond the details of cellular representations, there could be more focus on the nature of neural representations for flexibly representing all elements of an existing world. There could be inspiration from the framework used in computer animation. For example, the use of matrix implementations of two dimensional bezier surfaces (Sederberg 2012) or non-uniform rational beta-splines (NURBS) (Liu and Wang 2002) could be inspiration for how neurons generate the elements of an allocentric scene and the projection transformation into an egocentric view.

Words versus Equations: Verbal Hypotheses Versus Computational Models

This chapter has focused on a review of neural data and computational models. However, inspired by the interaction of neuroscientists and philosophers at the meeting in Tucson that prompted this edited volume, this section will venture out of the neuroscience expertise to discuss verbal hypotheses versus computational models.

The conference in Tucson was titled Time, Space and Memory, similar to elements of many other conferences and the title of this chapter. However, one could argue that neuroscience must eventually outgrow verbal terms such as memory and associated terms such as episodic memory, semantic memory, and working memory, and to develop more sophisticated representations for words such as space and time.

Science is ultimately a quest to create new symbolic languages of description that transcend prior verbal descriptions. Mature scientific fields such as Physics have a mathematical language of theories that transcend verbal descriptions. Words can be used to communicate and teach about physical theories, but they are not an essential component of the theory itself. The accepted scientific theories are consistent regardless of the inexact words or even variations in mathematical notation being used to describe it. Multiple theories ranging from Newtonian mechanics to the Schröodinger equations for atomic structures can be described entirely in equations that accurately guide experimental test. In most cases, the attempt to describe these theories in words alone usually introduces inaccuracies or misunderstandings. Similarly, the structure of molecules ranging from simple compounds to functional proteins to the genetic code in DNA can be described with a sequence of letters that need no words from natural language. The sequence and quantitative characteristics of each the sequence of amino acids in a protein and their structure and interaction with other proteins or small molecules can be functionally modelled, again without words.

At this point, systems neuroscience and psychology are still too immature as fields to have many accepted theories that transcend words. Theories at the single cell level can do this, using the mathematical framework of Rall and others to describe the membrane potential interactions across the structure of dendrites and axons, and using the Hodgkin-Huxley framework to describe the dynamics of voltage-sensitive and calcium-sensitive channels. Cellular neurophysiology is a mature science with theories that transcend words. However, systems neuroscience and behavioural neuroscience still suffer from the promiscuity of words to have multiple overlapping or nonoverlapping meanings.

Because of the immature nature of mathematical theory in systems neuroscience, words are necessary to describe experimental results for which there is not yet a quantitative theoretical framework. But sometimes scientific questions become overshadowed by disagreements over definitions of terms. As an alternative framework for an as yet unstructured field, the names of researchers can provide another type of symbolic representation for forming an interlocking web of collaborations and dates that give some structure to the broad mass of neuroscience data.

In the history of science, many domains of inquiry were initially domains of natural philosophy. Then, as empirical scientific inquiry grew more sophisticated within individual fields, the empirical components of individual fields of inquiry sequentially separated from philosophy and became independent disciplines. In some cases, the split seems to occur when the empirical data from experiments and the scientific theories about this data took a form that was independent from the verbal description of questions in the field. The philosophical questions remain in many areas of science, but the empirical theories in physics, chemistry and molecular biology have moved beyond verbal descriptions.

Replacing Words

How do we replace words in systems neuroscience? Based on examples from other fields such as physics, astronomy, chemistry, and engineering, the answer is to develop mathematical theories and notations that quantitatively account for the full range of data being addressed. What we currently inaccurately call episodic memory and semantic memory and working memory could eventually be addressed with a broad continuum model with multiple time scales and spatial scales that cannot be differentiated into these primitive terms. Or conversely, a theoretical framework might allow quantization of the dynamics of neural substrates that results in precise mathematical definitions of a specific number of memory elements that have no clear relationship to our current primitive terms. Terms like episodic, semantic, and working memory could be like the terms of air, water, fire, and earth used in early science, and could be replaced by an orderly periodic table of neural memory dynamical elements that predicts and guides quantitative research on memory (Hasselmo et al. 2021).

How do we define the structure of knowledge addressed in neuroscience? Neuroscience will benefit from a systematic theoretical structure that accurately bridges from the behavioural to the systems to the cellular and molecular levels (Levenstein et al. 2023). Unfortunately, this is a daunting task as the scope of neuroscience is not just the empirical data, but also the scope of knowledge represented in the human brain. These internal representations include everything that we can understand about the universe. In the simplest terms, the field addresses how the nervous system guides behaviour. We can focus on the central role of the nervous system, so we have a physically defined substrate. On the behavioural side, we could focus like a behaviourist on only the physical manifestations of behaviour—our movements in the world. However, this immediately expands into a description of everything, because we can utter words and write mathematical equations that attempt to describe our understanding of everything in the known universe. Thus the domain of neuroscience is ultimately the neural substrates for our human understanding of everything in the universe. In order to model how the nervous system guides behaviour, we need to model how the nervous system represents everything an individual can know about the universe, including an internal representation of the conscious self as a discrete entity interacting with the external world and planning future behaviour based on past memories (Hasselmo 2010). Ultimately, a model of memory function in cortical neural circuits must provide a framework for understanding how neurons can encode a memory representation of everything that a human can think about and remember, including the spatiotemporal trajectory of episodic memories. This is clearly a challenge, but an exciting challenge.

Acknowledgements

This work supported by the National Institutes of Health, grant numbers R01 MH120073, R01 MH60013, R01 MH052090 and by the Office of Naval Research MURI N00014-16-1-2832 and MURI N00014-19-1-2571 and DURIP N00014-17-1-2304. The authors have no conflict of interest.

References

Aggleton, J.P., Hunt, P.R., and Rawlins, J.N. (1986). The effects of hippocampal lesions upon spatial and non-spatial tests of working memory. *Behavioural Brain Research* 19(2), 133–46.

Aggleton, J.P., Neave, N., Nagle, S., and Hunt, P.R. (1995). A comparison of the effects of anterior thalamic, mamillary body and fornix lesions on reinforced spatial alternation. *Behav Brain Res* 68(1), 91–101.

Ainge, J.A., van der Meer, M.A., Langston, R.F., and Wood, E.R. (2007). Exploring the role of context-dependent hippocampal activity in spatial alternation behavior. *Hippocampus* 17(10), 988–1002.

Alexander, A.S., Carstensen, L.C., Hinman, J.R., Raudies, F., Chapman, G.W., and Hasselmo, M.E. (2020). Egocentric boundary vector tuning of the retrosplenial cortex. *Sci Adv* 6(8). eaaz2322.

Alexander, A.S., and Nitz, D.A. (2015). Retrosplenial cortex maps the conjunction of internal and external spaces. *Nature Neuroscience* 18, 1143–51.

Alexander, A.S., and Nitz, D.A. (2017). Spatially periodic activation patterns of retrosplenial cortex encode route sub-spaces and distance traveled. *Current Biology: CB* 27, 1551–60. e1554.

Alexander, A.S., Rangel, L.M., Tingley, D., and Nitz, D.A. (2018). Neurophysiological signatures of temporal coordination between retrosplenial cortex and the hippocampal formation. *Behav Neurosci* 132(5), 453–68.

Alexander, A.S., Robinson, J.C., Stern, C.E., and Hasselmo, M.E. (2023). Gated transformations from egocentric to allocentric reference frames involving retrosplenial cortex, entorhinal cortex, and hippocampus. *Hippocampus* 33(5), 465–87.

Arleo, A., Smeraldi, F., and Gerstner, W. (2004). Cognitive navigation based on nonuniform Gabor space sampling, unsupervised growing networks, and reinforcement learning. *IEEE Trans Neural Netw* 15(3), 639–52.

Banino, A., Barry, C., Uria, B., Blundell, C., Lillicrap, T., Mirowski, P., Pritzel, A. et al. (2018). Vector-based navigation using grid-like representations in artificial agents. *Nature* 557(7705), 429–33.

Barry, C., Hayman, R., Burgess, N., and Jeffery, K.J. (2007). Experience-dependent rescaling of entorhinal grids. *Nat Neurosci* 10(6), 682–84.

Barry, C., Lever, C., Hayman, R., Hartley, T., Burton, S., O'Keefe, J., Jeffery, K. et al. (2006). The boundary vector cell model of place cell firing and spatial memory. *Rev Neurosci* 17(1–2), 71–97.

Bicanski, A., and Burgess, N. (2018). A neural-level model of spatial memory and imagery. *Elife* 7, e33752.

Bittner, K.C., Grienberger, C., Vaidya, S.P., Milstein, A.D., Macklin, J.J., Suh, J., Tonegawa, S. et al. (2015). Conjunctive input processing drives feature selectivity in hippocampal CA1 neurons. *Nat Neurosci* 18(8), 1133–42.

Bittner, K.C., Milstein, A.D., Grienberger, C., Romani, S., and Magee, J.C. (2017). Behavioral time scale synaptic plasticity underlies CA1 place fields. *Science* 357(6355), 1033–36.

Boyle, A. (2021). Remembering events and representing time. *Synthese* 199, 2505–24.

Brandon, M.P., Bogaard, A.R., Libby, C.P., Connerney, M.A., Gupta, K., and Hasselmo, M.E. (2011). Reduction of theta rhythm dissociates grid cell spatial periodicity from directional tuning. *Science* 332(6029), 595–99.

Brandon, M.P., Bogaard, A.R., Schultheiss, N.W., and Hasselmo, M.E. (2013). Segregation of cortical head direction cell assemblies on alternating theta cycles. *Nat Neurosci* 16(6), 739–48.

Brandon, M.P., Koenig, J., Leutgeb, J.K., and Leutgeb, S. (2014). New and distinct hippocampal place codes are generated in a new environment during septal inactivation. *Neuron* 82(4), 789–96.

Bright, I.M., Meister, M.L.R., Cruzado, N.A., Tiganj, Z., Buffalo, E.A., and Howard, M.W. (2020). A temporal record of the past with a spectrum of time constants in the monkey entorhinal cortex. *Proc Natl Acad Sci U S A* 117(33), 20274–83.

Brioni, J.D., Decker, M.W., Gamboa, L.P., Izquierdo, I., and McGaugh, J.L. (1990). Muscimol injections in the medial septum impair spatial learning. *Brain Res* 522(2), 227–34.

Brown, E.N., Frank, L.M., Tang, D., Quirk, M.C., and Wilson, M.A. (1998). A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *J Neurosci* 18(18), 7411–25.

Brown, T.I., Hasselmo, M.E., and Stern, C.E. (2014). A high-resolution study of hippocampal and medial temporal lobe correlates of spatial context and prospective overlapping route memory. *Hippocampus* 24(7), 819–39.

Brown, T.I., Ross, R.S., Keller, J.B., Hasselmo, M.E., and Stern, C.E. (2010). Which way was I going? Contextual retrieval supports the disambiguation of well learned overlapping navigational routes. *J Neurosci* 30(21), 7414–22.

Brown, T.I., and Stern, C.E. (2014). Contributions of medial temporal lobe and striatal memory systems to learning and retrieving overlapping spatial memories. *Cereb Cortex* 24(7), 1906–22.

Brunec, I.K., and Momennejad, I. (2022). Predictive representations in hippocampal and prefrontal hierarchies. *J Neurosci* 42(2), 299–312.

Buetfering, C., Allen, K., and Monyer, H. (2014). Parvalbumin interneurons provide grid cell-driven recurrent inhibition in the medial entorhinal cortex. *Nat Neurosci* 17(5), 710–18.

Burgess, N. (2008). Grid cells and theta as oscillatory interference: Theory and predictions. *Hippocampus* 18(12), 1157–74.

Burgess, N., Barry, C., and O'Keefe, J. (2007). An oscillatory interference model of grid cell firing. *Hippocampus* 17(9), 801–12.

Burgess, N., Donnett, J.G., Jeffery, K.J., and O'Keefe, J. (1997). Robotic and neuronal simulation of the hippocampus and rat navigation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 352(1360), 1535–43.

Burgess, N., Jackson, A., Hartley, T., and O'Keefe, J. (2000). Predictions derived from modelling the hippocampal role in navigation. *Biol Cybern* 83(3), 301–12.

Buzsaki, G., and Tingley, D. (2018). Space and time: The hippocampus as a sequence generator. *Trends Cogn Sci* 22(10), 853–69.

Byrne, P., Becker, S., and Burgess, N. (2007). Remembering the past and imagining the future: a neural model of spatial memory and imagery. *Psychol Rev* 114(2), 340–75.

Cai, D.J., Aharoni, D., Shuman, T., Shobe, J., Biane, J., Song, W., Wei, B. et al. (2016). A shared neural ensemble links distinct contextual memories encoded close in time. *Nature* 534(7605), 115–18.

Carstensen, L.C., Alexander, A.S., Chapman, G.W., Lee, A.J., and Hasselmo, M.E. (2021). Neural responses in retrosplenial cortex associated with environmental alterations. *iScience* 24. 103377.

Cei, A., Girardeau, G., Drieu, C., Kanbi, K.E., and Zugaro, M. (2014). Reversed theta sequences of hippocampal cell assemblies during backward travel. *Nat Neurosci* 17(5), 719–24.

Chance, F.S. (2012). Hippocampal phase precession from dual input components. *J Neurosci* 32(47), 16693–703a.

Chen, G., Manson, D., Cacucci, F., and Wills, T.J. (2016). Absence of visual input results in the disruption of grid cell firing in the mouse. *Current Biology* 26(17), 2335–42.

Chen, S., Wang, J., and Siegelbaum, S.A. (2001). Properties of hyperpolarization-activated pacemaker current defined by coassembly of HCN1 and HCN2 subunits and basal modulation by cyclic nucleotide. *J Gen Physiol* 117(5), 491–504.

Chrastil, E.R., Sherrill, K.R., Hasselmo, M.E., and Stern, C.E. (2015). There and back again: Hippocampus and retrosplenial cortex track homing distance during human path integration. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 35(46), 15442–52.

Chrastil, E.R., Sherrill, K.R., Hasselmo, M.E., and Stern, C.E. (2016). Which way and how far? Tracking of translation and rotation information for human path integration. *Human Brain Mapping* 37, 3636–55.

Chrobak, J.J., Stackman, R.W., and Walsh, T.J. (1989). Intraseptal administration of muscimol produces dose-dependent memory impairments in the rat. *Behavioral and Neural Biology* 52, 357–69.

Clancy, K.B., Orsolic, I., and Mrsic-Flogel, T.D. (2019). Locomotion-dependent remapping of distributed cortical networks. *Nat Neurosci* 22(5), 778–86.

Climer, J.R., Newman, E.L., and Hasselmo, M.E. (2013). Phase coding by grid cells in unconstrained environments: Two-dimensional phase precession. *Eur J Neurosci* 38(4), 2526–41.

Conway, M.A. (2009). Episodic memories. *Neuropsychologia* 47(11), 2305–13.

Costa, V.C., Bueno, J.L., and Xavier, G.F. (2005). Dentate gyrus-selective colchicine lesion and performance in temporal and spatial tasks. *Behav Brain Res* 160(2), 286–303.

Dannenberg, H., Kelley, C., Hoyland, A., Monaghan, C.K., and Hasselmo, M.E. (2019). The firing rate speed code of entorhinal speed cells differs across behaviorally relevant time scales and does not depend on medial septum inputs. *J Neurosci* 39(18), 3434–53.

Dannenberg, H., Lazaro, H., Nambiar, P., Hoyland, A., and Hasselmo, M.E. (2020). Effects of visual inputs on neural dynamics for coding of location and running speed in medial entorhinal cortex. *Elife* 9. e62500.

Dayan, P. (1993). Improving generalization for temporal difference learning: The successor representation. *Neural Computation* 5(4), 613–24.

DeLuca, J., and Cicerone, K.D. (1991). Confabulation following aneurysm of the anterior communicating artery. *Cortex* 27(3), 417–23.

Deshmukh, S.S., and Knierim, J.J. (2011). Representation of non-spatial and spatial information in the lateral entorhinal cortex. *Front Behav Neurosci* 5, 69.

Deshmukh, S.S., and Knierim, J.J. (2013). Influence of local objects on hippocampal representations: Landmark vectors and memory. *Hippocampus*, 23(4), 253–267.

Do, Q., and Hasselmo, M.E. (2021). Neural circuits and symbolic processing. *Neurobiol Learn Mem* 186. 107552.

Douchamps, V., Jeewajee, A., Blundell, P., Burgess, N., and Lever, C. (2013). Evidence for encoding versus retrieval scheduling in the hippocampus by theta phase and acetylcholine. *J Neurosci* 33(20), 8689–704.

Eichenbaum, H. (2014). Time cells in the hippocampus: A new dimension for mapping memories. *Nat Rev Neurosci* 15(11), 732–44.

Eichenbaum, H., Dudchenko, P., Wood, E., Shapiro, M., and Tanila, H. (1999). The hippocampus, memory, and place cells: Is it spatial memory or a memory space? *Neuron* 23(2), 209–26.

Eichenbaum, H., Stewart, C., and Morris, R.G. (1990). Hippocampal representation in place learning. *J Neurosci* 10(11), 3531–42.

Eliasmith, C., Stewart, T.C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., and Rasmussen, D. (2012). A large-scale model of the functioning brain. *Science* 338(6111), 1202–05.

Emerich, D.F., and Walsh, T.J. (1989). Selective working memory impairments following intradentate injection of colchicine: Attenuation of the behavioral but not the neuropathological effects by gangliosides GM1 and AGF2. *Physiol Behav* 45(1), 93–101.

Erdem, U.M., and Hasselmo, M. (2012). A goal-directed spatial navigation model using forward trajectory planning based on grid cells. *Eur J Neurosci* 35(6), 916–31.

Erdem, U.M., and Hasselmo, M.E. (2014). A biologically inspired hierarchical goal directed navigation model. *J Physiol Paris* 108(1), 28–37.

Erdem, U.M., Milford, M.J., and Hasselmo, M.E. (2015). A hierarchical model of goal directed navigation selects trajectories in a visual environment. *Neurobiol Learn Mem* 117, 109–21.

Feng, T., Silva, D., and Foster, D.J. (2015). Dissociation between the experience-dependent development of hippocampal theta sequences and single-trial phase precession. *J Neurosci* 35(12), 4890–902.

Fenton, A.A., Kao, H.Y., Neymotin, S.A., Olypher, A., Vayntrub, Y., Lytton, W.W., and Ludvig, N. (2008). Unmasking the CA1 ensemble place code by exposures to small and large environments: More place cells and multiple, irregularly arranged, and expanded place fields in the larger space. *J Neurosci* 28(44), 11250–62.

Fenton, A.A., and Muller, R.U. (1998). Place cell discharge is extremely variable during individual passes of the rat through the firing field. *Proc. Natl. Acad. Sci. USA* 95(6), 3182–87.

Ferbinteanu, J., and Shapiro, M,L. (2003). Prospective and retrospective memory coding in the hippocampus. *Neuron* 40(6), 1227–39.

Foster, D.J., and Wilson, M.A. (2007). Hippocampal theta sequences. *Hippocampus* 17(11), 1093–99.

Frank, L.M., Brown, E.N., and Wilson, M. (2000). Trajectory encoding in the hippocampus and entorhinal cortex. *Neuron* 27(1), 169–78.

Giocomo, L.M., Stensola, T., Bonnevie, T., Van Cauter, T., Moser, M.B., and Moser, E.I. (2014). Topography of head direction cells in medial entorhinal cortex. *Curr Biol* 24(3), 252–62.

Gofman, X., Tocker, G., Weiss, S., Boccara, C.N., Lu, L., Moser, M.B., Moser, E.I. et al. (2019). Dissociation between postrhinal cortex and downstream parahippocampal regions in the representation of egocentric boundaries. *Curr Biol* 29(16), 2751–2757. e2754.

Goodridge, J.P., and Taube, J.S. (1997). Interaction between the postsubiculum and anterior thalamus in the generation of head direction cell activity. *J Neurosci* 17(23), 9315–30.

Green, E.J., and Schellenberg, S. (2017). Spatial perception: The perspectival aspect of perception. *Philosophy Compass* 13(2). e12472.

Grienberger, C., Milstein, A.D., Bittner, K.C., Romani, S., and Magee, J.C. (2017). Inhibitory suppression of heterogeneously tuned excitation enhances spatial coding in CA1 place cells. *Nat Neurosci* 20(3), 417–26.

Griffin, A.L., Eichenbaum, H., and Hasselmo, M.E. (2007). Spatial representations of hippocampal CA1 neurons are modulated by behavioral context in a hippocampus-dependent memory task. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 27(9), 2416–23.

Hafting, T., Fyhn, M., Bonnevie, T., Moser, M.B., and Moser, E.I. (2008). Hippocampus-independent phase precession in entorhinal grid cells. *Nature* 453(7199), 1248–52.

Hafting, T., Fyhn, M., Molden, S., Moser, M.-B., and Moser, E.I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature* 436, 801–06.

Hallock, H.L., Arreola, A.C., Shaw, C.L., and Griffin, A.L. (2013). Dissociable roles of the dorsal striatum and dorsal hippocampus in conditional discrimination and spatial alternation T-maze tasks. *Neurobiology of Learning and Memory* 100, 108–16.

Hartley, T., Burgess, N., Lever, C., Cacucci, F., and O'Keefe, J. (2000). Modeling place fields in terms of the cortical inputs to the hippocampus. *Hippocampus* 10(4), 369–79.

Hartley, T., Lever, C., Burgess, N., and O'Keefe, J. (2014). Space in the brain: How the hippocampal formation supports spatial cognition. *Philos Trans R Soc Lond B Biol Sci* 369(1635). 20120510.

Hasselmo, M.E. (2005). What is the function of hippocampal theta rhythm?—Linking behavioral data to phasic properties of field potential and unit recording data. *Hippocampus* 15(7), 936–49.

Hasselmo, M.E. (2006). The role of acetylcholine in learning and memory. *Curr Opin Neurobiol* 16(6), 710–15.

Hasselmo, M.E. (2008). Grid cell mechanisms and function: Contributions of entorhinal persistent spiking and phase resetting. *Hippocampus* 18(12), 1213–29.

Hasselmo, M.E. (2009). A model of episodic memory: Mental time travel along encoded trajectories using grid cells. *Neurobiol Learn Mem* 92(4), 559–73.

Hasselmo, M.E. (2010). Consciousness and neural time travel. In E. Perry, D. Collerton, F.E. LeBeau et al. (eds). *New Horizons in the Neuroscience of Consciousness* (pp. 73–80). Philadelphia, PA: John Benhamins Publishing Company.

Hasselmo, M.E. (2012). *How We Remember: Brain Mechanisms of Episodic Memory*. Cambridge, MA: MIT Press.

Hasselmo, M.E. (2015). Remembering by index and content: Response to Sarah Robins. *Philosophical Psychology* 26(8), 916–19.

Hasselmo, M.E., Alexander, A.S., Hoyland, A., Robinson, J.C., Bezaire, M.J., Chapman, G.W., Saudargiene, A. et al. (2021). The unexplored territory of neural models: potential guides for exploring the function of metabotropic neuromodulation. *Neuroscience* 456, 143–58.

Hasselmo, M.E., Bodelón, C., and Wyble, B.P. (2002). A proposed function for hippocampal theta rhythm: Separate phases of encoding and retrieval enhance reversal of prior learning. *Neural Computation* 14(4), 793–817.

Hasselmo, M.E., and Eichenbaum, H. (2005). Hippocampal mechanisms for the context-dependent retrieval of episodes. *Neural Netw* 18(9), 1172–90.

Hasselmo, M.E., Giocomo, L.M., Brandon, M.P., and Yoshida, M. (2010). Cellular dynamical mechanisms for encoding the time and place of events along spatiotemporal trajectories in episodic memory. *Behav Brain Res* 215(2), 261–74.

Hasselmo, M.E., and Wyble, B.P. (1997). Free recall and recognition in a network model of the hippocampus: Simulating effects of scopolamine on human memory function. *Behav Brain Res* 89(1–2), 1–34.

Heys, J.G., and Dombeck, D.A. (2018). Evidence for a subcircuit in medial entorhinal cortex representing elapsed time during immobility. *Nat Neurosci* 21(11), 1574–82.

Heys, J.G., Rangarajan, K.V., and Dombeck, D.A. (2014). The functional micro-organization of grid cells revealed by cellular-resolution imaging. *Neuron* 84(5), 1079–90.

Hinman, J.R., Brandon, M.P., Climer, J.R., Chapman, G.W., and Hasselmo, M.E. (2016). Multiple running speed signals in medial entorhinal cortex. *Neuron* 91(3), 666–79.

Hinman, J.R., Chapman, G.W., and Hasselmo, M.E. (2019). Neuronal representation of environmental boundaries in egocentric coordinates. *Nat Commun* 10(1), 2772.

Howard, M.W., and Eichenbaum, H. (2013). The hippocampus, time, and memory across scales. *J Exp Psychol Gen* 142(4), 1211–30.

Howard, M.W., MacDonald, C.J., Tiganj, Z., Shankar, K.H., Du, Q., Hasselmo, M.E., and Eichenbaum, H. (2014). A unified mathematical framework for coding time, space, and sequences in the hippocampal region. *J Neurosci* 34(13), 4692–707.

Hoydal, O.A., Skytoen, E.R., Andersson, S.O., Moser, M.B., and Moser, E.I. (2019). Object-vector coding in the medial entorhinal cortex. *Nature* 568(7752), 400–04.

Huxter, J., Burgess, N., and O'Keefe, J. (2003). Independent rate and temporal coding in hippocampal pyramidal cells. *Nature* 425(6960), 828–32.

Huxter, J.R., Senior, T.J., Allen, K., and Csicsvari, J. (2008). Theta phase-specific codes for two-dimensional position, trajectory and heading in the hippocampus. *Nat Neurosci* 11(5), 587–94.

Jackson, J., and Redish, A.D. (2007). Network dynamics of hippocampal cell-assemblies resemble multiple spatial maps within single tasks. *Hippocampus* 17(12), 1209–29.

Jeewajee, A., Barry, C., Douchamps, V., Manson, D., Lever, C., and Burgess, N. (2014). Theta phase precession of grid and place cell firing in open environments. *Philos Trans R Soc Lond B Biol Sci* 369(1635). 20120532.

Jeewajee, A., Barry, C., O'Keefe, J., and Burgess, N. (2008). Grid cells and theta as oscillatory interference: Electrophysiological data from freely moving rats. *Hippocampus* 18(12), 1175–85.

Jeffery, K.J., Donnett, J.G., and O'Keefe, J. (1995). Medial septal control of theta-correlated unit firing in the entorhinal cortex of awake rats. *Neuroreport* 6(16), 2166–70.

Johnson, A., and Redish, A.D. (2007). Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *J Neurosci* 27(45), 12176–89.

Kay, K., Chung, J.E., Sosa, M., Schor, J.S., Karlsson, M.P., Larkin, M.C., Liu, D.F. et al. (2020). Constant sub-second cycling between representations of possible futures in the hippocampus. *Cell* 180(3), 552–567. e525.

Kelley, C., Dura-Bernal, S., Neymotin, S.A., Antic, S.D., Carnevale, N.T., Migliore, M., and Lytton, W.W. (2021). Effects of I(h). and TASK-like shunting current on dendritic impedance in layer 5 pyramidal-tract neurons. *J Neurophysiol* 125(4), 1501–16.

Kinsky, N.R., Mau, W., Sullivan, D.W., Levy, S.J., Ruesch, E.A., and Hasselmo, M.E. (2020). Trajectory-modulated hippocampal neurons persist throughout memory-guided navigation. *Nat Commun* 11(1), 2443.

Knierim, J.J., Kudrimoti, H.S., and McNaughton, B.L. (1998). Interactions between idiothetic cues and external landmarks in the control of place cells and head direction cells. *J Neurophysiol* 80(1), 425–46.

Koenig, J., Linder, A.N., Leutgeb, J.K., and Leutgeb, S. (2011). The spatial periodicity of grid cells is not sustained during reduced theta oscillations. *Science* 332(6029), 592–95.

Kraus, B.J., Brandon, M.P., Robinson, R.J., 2nd, Connerney, M.A., Hasselmo, M.E., and Eichenbaum, H. (2015). During running in place, grid cells integrate elapsed time and distance run. *Neuron* 88(3), 578–89.

Kraus, B.J., Robinson, R.J., 2nd, White, J.A., Eichenbaum, H., and Hasselmo, M.E. (2013). Hippocampal 'time cells': Time versus path integration. *Neuron* 78(6), 1090–101.

Kropff, E., Carmichael, J.E., Moser, M.B., and Moser, E.I. (2015). Speed cells in the medial entorhinal cortex. *Nature* 523(7561), 419–24.

Kubie, J.L. and Fenton, A.A. (2012). Linear look-ahead in conjunctive cells: An entorhinal mechanism for vector-based navigation. *Front Neural Circuits* 6, 20.

LaChance, P.A., and Taube, J.S. (2023). A model for transforming egocentric views into goal-directed behavior. *Hippocampus* 33(5), 488–504.

LaChance, P.A., Todd, T.P., and Taube, J.S. (2019). A sense of space in postrhinal cortex. *Science* 365(6449). eaax4192.

Levenstein, D., Alvarez, V.A., Amarasingham, A., Azab, H., Chen, Z.S., Gerkin, R.C., Hasenstaub, A. et al. (2023). On the role of theory and modeling in neuroscience. *J Neurosci* 43(7), 1074–88.

Lever, C., Burton, S., Jeewajee, A., O'Keefe, J., and Burgess, N. (2009). Boundary vector cells in the subiculum of the hippocampal formation. *The Journal of Neuroscience* 29, 9771–77.

Lever, C., Wills, T., Cacucci, F., Burgess, N., and O'Keefe, J. (2002). Long-term plasticity in hippocampal place-cell representation of environmental geometry. *Nature* 416(6876), 90–94.

Levy, S.J., Kinsky, N.R., Mau, W., Sullivan, D.W., and Hasselmo, M.E. (2021). Hippocampal spatial memory representations in mice are heterogeneously stable. *Hippocampus* 31(3), 244–60.

Levy, W.B. (1996). A sequence predicting CA3 is a flexible associator that learns and uses context to solve hippocampal-like tasks. *Hippocampus* 6(6), 579–90.

Lipton, P.A., White, J.A., and Eichenbaum, H. (2007). Disambiguation of overlapping experiences by neurons in the medial entorhinal cortex. *J Neurosci* 27(21), 5787–95.

Lisman, J., and Redish, A.D. (2009). Prediction, sequences and the hippocampus. *Philos Trans R Soc Lond B Biol Sci* 364(1521), 1193–201.

Liu, L., and Wang, G. (2002). Explicit matrix representation for NURBS curves and surfaces. *Computer Aided Geometric Design* 19, 409–19.

Liu, Y., Levy, S., Mau, W., Geva, N., Rubin, A., Ziv, Y., Hasselmo, M. et al. (2022). Consistent population activity on the scale of minutes in the mouse hippocampus. *Hippocampus* 32(5), 359–72.

Liu, Y., Tiganj, Z., Hasselmo, M.E., and Howard, M.W. (2019). A neural microcircuit model for a scalable scale-invariant representation of time. *Hippocampus* 29(3), 260–74.

MacDonald, C.J., Carrow, S., Place, R., and Eichenbaum, H. (2013). Distinct hippocampal time cell sequences represent odor memories in immobilized rats. *J Neurosci* 33(36), 14607–16.

MacDonald, C.J., Lepage, K.Q., Eden, U.T., and Eichenbaum, H. (2011). Hippocampal 'time cells' bridge the gap in memory for discontiguous events. *Neuron* 71(4), 737–49.

Magee, J.C. (1999). Dendritic lh normalizes temporal summation in hippocampal CA1 neurons. *Nat Neurosci* 2(6), 508–14.

Magee, J.C. (2001). Dendritic mechanisms of phase precession in hippocampal CA1 pyramidal neurons. *J Neurophysiol* 86(1), 528–32.

Mankin, E.A., Diehl, G.W., Sparks, F.T., Leutgeb, S., and Leutgeb, J.K. (2015). Hippocampal CA2 activity patterns change over time to a larger extent than between spatial contexts. *Neuron* 85(1), 190–201.

Manns, J.R., Zilli, E.A., Ong, K.C., Hasselmo, M.E., and Eichenbaum, H. (2007). Hippocampal CA1 spiking during encoding and retrieval: Relation to theta phase. *Neurobiol Learn Mem* 87(1), 9–20.

Mau, W., Sullivan, D.W., Kinsky, N.R., Hasselmo, M.E., Howard, M.W., and Eichenbaum, H. (2018). The same hippocampal ca1 population simultaneously codes temporal information over multiple timescales. *Current Biology* 28(10), 1499–1508. e1494.

Maurer, A.P., Cowen, S.L., Burke, S.N., Barnes, C.A., and McNaughton, B.L. (2006). Phase precession in hippocampal interneurons showing strong functional coupling to individual pyramidal cells. *J Neurosci* 26(52), 13485–92.

Maurer, A.P., Lester, A.W., Burke, S.N., Ferng, J.J., and Barnes, C.A. (2014). Back to the future: preserved hippocampal network activity during reverse ambulation. *J Neurosci* 34(45), 15022–31.

McNaughton, B.L., Barnes, C.A., and O'Keefe, J. (1983). The contributions of position, direction, and velocity to single unit-activity in the hippocampus of freely-moving rats. *Experimental Brain Research* 52(1), 41–49.

McNaughton, B.L., Mizumori, S.J., Barnes, C.A., Leonard, B.J., Marquis, M., and Green, E.J. (1994). Cortical representation of motion during unrestrained spatial navigation in the rat. *Cereb Cortex* 4(1), 27–39.

McNaughton, B.L., and Morris, R.G.M. (1987). Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends Neurosci* 10, 408–15.

Mitchell, S.J., Rawlins, J.N., Steward, O., and Olton, D.S. (1982). Medial septal area lesions disrupt theta rhythm and cholinergic staining in medial entorhinal cortex and produce impaired radial arm maze behavior in rats. *J Neurosci* 2(3), 292–302.

Momennejad, I. (2020). Learning structures: Predictive representations, replay, and generalization. *Curr Opin Behav Sci* 32, 155–66.

Morris, R.G., Garrud, P., Rawlins, J.N., and O'Keefe, J. (1982). Place navigation impaired in rats with hippocampal lesions. *Nature* 297(5868), 681–83.

Muller, R.U., and Kubie, J.L. (1987). The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. *J Neurosci* 7(7), 1951–68.

Muller, R.U., Kubie, J.L., and Ranck, J.B., Jr. (1987). Spatial firing patterns of hippocampal complex-spike cells in a fixed environment. *J Neurosci* 7(7), 1935–50.

Munn, R.G.K., Mallory, C.S., Hardcastle, K., Chetkovich, D.M., and Giocomo, L.M. (2020). Entorhinal velocity signals reflect environmental geometry. *Nat Neurosci* 23(2), 239–51.

Murdock, B. (2005). Storage and retrieval of serial-order information. *Memory* 13(3–4), 259–66.

Naud, R., and Sprekeler, H. (2018). Sparse bursts optimize information transmission in a multiplexed neural code. *Proc Natl Acad Sci U S A* 115(27), E6329–E6338.

Nelson A., and Mooney R. (2016). The basal forebrain and motor cortex provide convergent yet distinct movement-related inputs to the auditory cortex. *Neuron* 90(3), 635–48.

Niell, C.M., and Stryker, M.P. (2010). Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron* 65, 472–79.

Ning, W., Bladon, J.H., and Hasselmo, M.E. (2022). Complementary representations of time in the prefrontal cortex and hippocampus. *Hippocampus* 32(8), 577–96.

O'Keefe, J. (1976). Place units in the hippocampus of the freely moving rat. *Exp Neurol* 51(1), 78–109.

O'Keefe, J. (1990). A computational theory of the hippocampal cognitive map. *Prog. Brain Res* 83, 301–12.

O'Keefe, J., and Burgess, N. (1996). Geometric determinants of the place fields of hippocampal neurons. *Nature* 381, 425–28.

O'Keefe, J., Burgess. N., Donnett, J.G., Jeffery, K.J., and Maguire, E.A. (1998). Place cells, navigational accuracy, and the human hippocampus. *Philos Trans R Soc Lond B Biol Sci* 353(1373), 1333–40.

O'Keefe, J., and Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res* 34, 171–75.

O'Keefe, J., and Nadel, L. (1978). *The Hippocampus as a Cognitive Map*. Oxford, New York: Oxford University Press.

O'Keefe, J., Nadel, L., Keightley, S., and Kill, D. (1975). Fornix lesions selectively abolish place learning in the rat. *Exp Neurol* 48(1), 152–66.

O'Keefe, J., and Recce, M.L. (1993). Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* 3, 317–30.

O'Neill, J., Boccara, C.N., Stella, F., Schoenenberger, P., and Csicsvari, J. (2017). Superficial layers of the medial entorhinal cortex replay independently of the hippocampus. *Science* 355(6321), 184–88.

O'Reilly, R.C., and Munakata, Y. (2000). *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain.* Cambridge, MA: MIT Press.

Ormond, J., and O'Keefe, J. (2022). Hippocampal place cells have goal-oriented vector fields during navigation. *Nature* 607(7920), 741–46.

Pastalkova, E., Itskov, V., Amarasingham, A., and Buzsaki, G. (2008). Internally generated cell assembly sequences in the rat hippocampus. *Science* 321(5894), 1322–27.

Pérez-Escobar, J.A., Kornienko, O., Latuske, P., Kohler, L., and Allen, K. (2016). Visual landmarks sharpen grid cell metric and confer context specificity to neurons of the medial entorhinal cortex. *Elife* 5. e16937.

Phillips, I. (2014). Experience of and in time. *Philosophy Compass* 9(2), 131–44.

Plate, T.A. (1995). Holographic reduced representations. *IEEE Trans Neural Netw* 6(3), 623–41.

Poulter, S., Lee, S.A., Dachtler, J., Willis, T.J., and Lever, C. (2021). Vector trace cells in the subiculum of the hippocampal formation. *Nat. Neurosci.* 24(2), 266–75.

Raudies, F., Brandon, M.P., Chapman, G.W., and Hasselmo, M.E. (2015). Head direction is coded more strongly than movement direction in a population of entorhinal neurons. *Brain Res* 1621, 355–67.

Raudies, F., and Hasselmo, M.E. (2015). Differences in visual-spatial input may underlie different compression properties of firing fields for grid cell modules in medial entorhinal cortex. *PLoS Comput Biol* 11(11). e1004596.

Rawlins, J.N., Feldon, J., and Gray, J.A. (1979). Septo-hippocampal connections and the hippocampal theta rhythm. *Exp Brain Res* 37(1), 49–63.

Redish, A.D., and Touretzky, D.S. (1998). The role of the hippocampus in solving the Morris water maze. *Neural Comput.* 10(1), 73–111.

Redish, D. (1999). *Beyond the Cognitive Map: From Place Cells to Episodic Memory*.

Rigotti, M., Barak, O., Warden, M.R., Wang, X.-J., Daw, N.D., Miller, E.K., and Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature* 497, 585–90.

Robins, S.K. (2015). A mechanism for mental time travel? A critical review of Michael Hasselmo's *How We Remember: Brain Mechanisms of Episodic Memory*. *Philosophical Psychology* 28, 903–15.

Robinson, J.C., Brandon, M.P., and Hasselmo, M.E. (2019). Identifying the contribution of medial septum cell types in generating the entorhinal grid cell code. *Soc. Neurosci. Abstr.* 45, 164–102.

Rubin, A., Geva, N., Sheintuch, L., and Ziv, Y. (2015). Hippocampal ensemble dynamics timestamp events in long-term memory. *Elife* 4. e12247.

Rule, M.E., O'Leary, T., and Harvey, C.D. (2019). Causes and consequences of representational drift. *Curr. Opin. Neurobiol.* 58, 141–47.

Salz, D.M., Tiganj, Z., Khasnabish, S., Kohley, A., Sheehan, D., Howard, M.W., and Eichenbaum, H. (2016). Time Cells in Hippocampal Area CA3. *J Neurosci* 36(28), 7476–84.

Sargolini, F., Fyhn, M., Hafting, T., McNaughton, B.L., Witter, M.P., Moser, M.B., and Moser, E.I. (2006). Conjunctive representation of position, direction, and velocity in entorhinal cortex. *Science* 312(5774), 758–62.

Savelli, F., Yoganarasimha, D., and Knierim, J.J. (2008). Influence of boundary removal on the spatial representations of the medial entorhinal cortex. *Hippocampus* 18(12), 1270–82.

Schmidt, R., Diba, K., Leibold, C., Schmitz, D., Buzsaki, G., and Kempter, R. (2009). Single-trial phase precession in the hippocampus. *J Neurosci* 29(42), 13232–41.

Sederberg, T.W. (2012). Computer Aided Geometric Design. https://scholarsarchive.byu.edu/facpub/1: BYU ScholarsArchive.

Sharma, S., Chandra, S., and Fiete, I.R. (2022). Content-addressable memory without catastrophic forgetting by heteroassociation with a fixed scaffold. *Proceedings of the 39th International Conference on Machine Learning*. Baltimore, Maryland, USA: PLMR.

Sherrill, K.R., Chrastil, E.R., Ross, R.S., Erdem, U.M., Hasselmo, M.E., and Stern. C.E. (2015). Functional connections between optic flow areas and navigationally responsive brain regions during goal-directed navigation. *Neuroimage* 118, 386–96.

Sherrill, K.R., Erdem, U.M., Ross, R.S., Brown, T.I., Hasselmo, M.E., and Stern. C.E. (2013). Hippocampus and retrosplenial cortex combine path integration signals for successful navigation. *J Neurosci* 33(49), 19304–13.

Sheynikhovich, D., Chavarriaga, R., Strosslin, T., Arleo, A., and Gerstner, W. (2009). Is there a geometric module for spatial orientation? Insights from a rodent navigation model. *Psychol Rev* 116(3), 540–66.

Skaggs, W.E., McNaughton, B.L., Wilson, M.A., and Barnes, C.A. (1996). Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus* 6(2), 149–72.

Solstad, T., Boccara, C.N., Kropff, E., Moser, M.B., and Moser, E.I. (2008). Representation of geometric borders in the entorhinal cortex. *Science* 322(5909), 1865–68.

Stachenfeld, K.L., Botvinick, M.M., and Gershman, S.J. (2017). The hippocampus as a predictive map. *Nat. Neurosci.* 20(11), 1643–53.

Steffenach, H.A., Witter, M., Moser, M.B., and Moser, E.I. (2005). Spatial memory in the rat requires the dorsolateral band of the entorhinal cortex. *Neuron* 45(2), 301–13.

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

Sutton, N.M., and Ascoli, G.A. (2021). Spiking neural networks and hippocampal function: A web-accessible survey of simulations, modeling methods, and underlying theories. *Cogn Syst Res* 70, 80–92.

Taube, J.S. (1995). Head direction cells recorded in the anterior thalamic nuclei of freely moving rats. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 15(1), 70–86.

Taube, J.S., Kesslak, J.P., and Cotman C.W. (1992). Lesions of the rat postsubiculum impair performance on spatial tasks. *Behav Neural Biol* 57(2), 131–43.

Taube, J.S., Muller, R.U., and Ranck, J.B. (1990). Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 10(2), 420–35.

Terada, S., Sakurai, Y., Nakahara, H., and Fujisawa, S. (2017). Temporal and rate coding for discrete event sequences in the hippocampus. *Neuron* 94(6), 1248–62. e1244.

Tiganj, Z., Jung, M.W., Kim, J., and Howard, M.W. (2017). Sequential firing codes for time in rodent medial prefrontal cortex. *Cerebral Cortex* 27, 5663–71.

Touretzky, D.S., and Redish, A.D. (1996). Theory of rodent navigation based on interacting representations of space. *Hippocampus* 6(3), 247–70.

Traub, R., Miles, R., and Buzsaki, G. (1992). Computer simulation of carbachol-driven rhythmic population oscillations in the CA3 region of the in vitro rat hippocampus. *J. Physiol.* 451, 653–72.

Traub, R.D., Whittington, M.A., and Cunningham, M.O. (2022). Simulation of oscillatory dynamics induced by an approximation of grid cell output. *Reviews in the Neurosciences*, 34(5), 517–532.

Tsao, A., Sugar, J., Lu, L., Wang, C., Knierim, J.J., Moser, M.B., and Moser, E.I. (2018). Integrating time from experience in the lateral entorhinal cortex. *Nature* 561(7721), 57–62.

Tulving, E. (1984). Precis of Elements of episodic memory. *Behav Brain Sci* 7, 223–68.

Vaidya, S.P., and Johnston, D. (2013). Temporal synchrony and gamma-to-theta power conversion in the dendrites of CA1 pyramidal neurons. *Nature Neuroscience* 16, 1812–20.

vanWijngaarden, J.B., Babl, S.S., and Ito, H.T. (2020). Entorhinal-retrosplenial circuits for allocentric-egocentric transformation of boundary coding. *Elife* 9. e59816.

Wallenstein, G.V., and Hasselmo, M.E. (1997). GABAergic modulation of hippocampal population activity: sequence learning, place field development, and the phase precession effect. *J Neurophysiol* 78(1), 393–408.

Wang, C., Chen, X., and Knierim, J.J. (2020). Egocentric and allocentric representations of space in the rodent brain. *Curr. Opin. Neurobiol.* 60, 12–20.

Wang, C., Chen, X., Lee, H., Deshmukh, S.S., Yoganarasimha, D., Savelli, F., and Knierim, J.J. (2018). Egocentric coding of external items in the lateral entorhinal cortex. *Science* 362(6417), 945–49.

Wang, Y., Romani, S., Lustig, B., Leonardo, A., and Pastalkova, E. (2015). Theta sequences are essential for internally generated hippocampal firing fields. *Nat Neurosci* 18(2), 282–88.

Welday, A.C., Shlifer, I.G., Bloom, M.L., Zhang, K., and Blair, H.T. (2011). Cosine directional tuning of theta cell burst frequencies: Evidence for spatial coding by oscillatory interference. *J Neurosci* 31(45), 16157–76.

Wells, C.E., Amos, D.P., Jeewajee, A., Douchamps, V., Rodgers, J., O'Keefe, J., Burgess, N. et al. (2013). Novelty and anxiolytic drugs dissociate two components of hippocampal theta in behaving rats. *J Neurosci* 33(20), 8650–67.

Wikenheiser, A.M., and Redish, A.D. (2015). Hippocampal theta sequences reflect current goals. *Nat Neurosci* 18(2), 289–94.

Wills, T.J., Barry, C., and Cacucci, F. (2012). The abrupt development of adult-like grid cell firing in the medial entorhinal cortex. *Front Neural Circuits* 6, 21.

Wilmerding, L.K., Kondratyev, I., Ramirez, S., and Hasselmo, M.E. (2023). Route-dependent spatial engram tagging in mouse dentate gyrus. *Neurobiol Learn Mem* 200. 107738.

Winter, S.S., Clark, B.J., and Taube, J.S. (2015). Disruption of the head direction cell network impairs the parahippocampal grid cell signal. *Science* 347(6224), 870–74.

Wood, E.R., Dudchenko, P.A., Robitsek, R.J., and Eichenbaum, H. (2000). Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron* 27(3), 623–33.

Ziv, Y., Burns, L.D., Cocker, E.D., Hamel, E.O., Ghosh, K.K., Kitch, L.J., Gamal, A.E. et al. (2013). Long-term dynamics of CA1 hippocampal place codes. *Nat. Neurosci.* 16(3), 264–66.

Zugaro, M.B., Monconduit, L., and Buzsaki, G. (2005). Spike phase precession persists after transient intrahippocampal perturbation. *Nat Neurosci* 8(1), 67–71.

Zutshi, I., Leutgeb, J.K., and Leutgeb, S. (2017). Theta sequences of grid cell populations can provide a movement-direction signal. *Curr Opin Behav Sci* 17, 147–54.

9

Simulationism and Memory Traces

Felipe De Brigard

Introduction

I started graduate school almost twenty years ago, as a student in both philosophy as well as psychology and neuroscience. As a result, I was exposed to two very different views on the nature of memory. On the one hand, there was the philosopher's view, according to which the primary function of memory is to reproduce a past event or experience. Such reproduction—this view holds—is underwritten by the preservation of the encoded content in a memory trace, which is later retrieved in the act of remembering. Moreover, memory is thought to be essentially distinct from imagination, not only because remembering, unlike imagining, requires a causal connection to the remembered event, but also because the verb 'remembering' is thought to be factive. That is, for one to truthfully express the proposition 'S remembers that p', then it must be the case that p obtained. One cannot remember what didn't happen; at best, one can imagine it. Sure, one may experience from time to time mental contents that do not correspond to actual events in one's past, or that perhaps distort them a bit, but such cases of false and distorted memories are, according to the philosopher's view, instances in which memory malfunctions (Kurtzman 1983).

On the other hand, there was the psychologist's view of memory. According to this view, memory isn't reproductive but reconstructive. Memories are encoded, not as individual events, but as instances of larger knowledge structures one has acquired through life (Bartlett 1932). As such, every act of encoding is embedded within an act of retrieval of related information that interprets and transforms the experienced content. There is also loss of information; because perception is fast and often incomplete, and because attention is limited and filtering, not everything we experience is encoded. And, of course, there is forgetting. Memory traces, if they exist at all, likely don't preserve all of the encoded content but, at best, an incomplete shadow of a past experience. Remembering is thus a reconstructive process in which past experiences are put back together by the joint operation of an incomplete memory trace and an active imagination that helps to fill the gaps at retrieval. It is because of this essential interaction between memory and imagination that false and distorted memories occasionally occur, although they do not reflect a failure in the system, but rather a natural by-product of the reconstructive operations of our mnemonic processes. 'Remembering', thus, need not be factive, for it is natural

Felipe De Brigard, *Simulationism and Memory Traces*. In: *Space, Time, and Memory*. Edited by: Lynn Nadel and Sara Aronowitz, Oxford University Press. DOI: 10.1093/oso/9780192882547.003.0009

to remember events that did not happen or that, when brought to mind, differ in subtle ways from how they actually occurred.

This tension between the preservationist view of the philosopher and the reconstructivist view of the psychologist is reflected today in two of the leading theories in the philosophy of memory: *causalism* and *simulationism*. The views are thought to be diametrically opposed and, as it happens in philosophy, practitioners tend to take sides. I did, too, about a decade ago (De Brigard 2014b). Yet, time has passed, arguments and counterarguments have been published, and new—and old—empirical findings have surfaced. It behooves us then to act wisely and to proportion our beliefs to the available evidence (Hume 1748/1977). And I think the evidence suggests that simulationism, in its original formulation, is wrong, and that it needs to be reformulated. Moreover, contra the original formulation of simulationism, I think we need to appeal to memory traces to explain remembering. The good news, though, is that reformulating simulationism will allow us to see the dichotomy between causalism and simulationism as a false one, opening thus the door for a reconciliation. To that end, I will start, in section 2, with a brief recount of both the motivations for, and standard formulations of, both causalism and simulationism. Next, in section 3, I will argue that two assumptions of the simulationist view are likely false and, thus, that the theory needs to be updated. Finally, in section 4, I show how this updated version of simulationism can help to dispel the false dichotomy between causalism and simulationism.

Causalism and Simulationism

Although the idea that memories are causally linked to the past events they are about can be traced as far back as Aristotle (De Brigard 2023), causalism in its contemporary form is typically associated to the celebrated paper 'Remembering' by Martin and Deutscher (1966). I like to think of the original motivation behind causalism as threefold. First, there was a patent dissatisfaction with classical representationalist views of memory, as they tended to advocate for internal or subjective criteria to distinguish memories from imaginations. Both rationalist and empiricist philosophers held that memories were preserved ideas of past perceptions or experiences, and that they could be distinguished from imaginations thanks to some internal or subjective criterion, such as vivacity, familiarity, coherence with other beliefs, or even apperceptions (De Brigard 2019). Such 'memory markers', unfortunately, are problematic, as it was easy to come up with counterexamples showing that neither of them constitutes necessary nor sufficient conditions for a present mental representation to count as a memory, as opposed to a perception or an imagination. A second motivation stem from an increasing dissatisfaction with direct realism, the view according to which remembering involved no intermediate representations, but rather a particular kind of direct acquaintance with the remembered past event. Alas, the mysterious nature of this direct acquaintance

relation with the past was metaphysically suspicious, and by the mid-twentieth century direct realism was all but discredited (Furlong 1948). Finally, there was also discontent with the behaviourist alternative which, driven by Wittgenstein's criticisms against the use of mental representation to explain psychological phenomena, inspired Malcolm to reject the claim that a proper account of remembering required talk of memory traces or any causal connection with the remembered event at all (Malcolm 1963).

However, as Martin and Deutscher (1966) neatly show, the need for a causal connection to the remembered event becomes indispensable when we try to distinguish cases of actual remembering from cases of apparent imagining, apparent remembering, and relearning (Robins 2016; Michaelian and Robins 2018). Cases of apparent imagining involve individuals that bring to mind a mental content they think they are conjuring up anew, but turn out to be recollections of past events whose acquisition they had forgotten about. Cases of apparent remembering involve individuals entertaining mental contents they take to be recollections but in fact correspond to no past experiences at all. And, finally, cases of relearning involve individuals that have encoded particular contents, then forget about them, and then relearn them from a different source or via some deviant causal chain. Thus, to avoid these kinds of cases, and secure the need for a causal condition, Martin and Deutscher (1966) offered a causal view of remembering, according to which an individual, *S*, remembers a past event, *e*, if and only if:

1) *S* now, at retrieval, represents *e* [*Current representation*]
2) *S* represented *e* at the time of encoding [*Past representation*], and
3) There is an appropriate causal connection between the content represented at encoding and the content represented at retrieval [*Appropriate causation*]

The qualifier 'appropriate' here is critical for the view, as it allows it to rule out cases of remembering that occur due to deviant causal chains or serendipitous relearning. Moreover, the appropriateness of the causal connection is in turn safeguarded by the existence of a stored memory trace, representing the content formed at encoding, and recovered, unchanged, at retrieval. Note that although there are some variations among causalists—some, for instance allow certain differences between the encoded and retrieved event (Bernecker 2009)—all of them accept the necessity of the causal connection between the experienced and the remembered event.

However, in the past decade, the necessity of an appropriate causal connection between the experienced and the remembered event has been questioned by simulationism, a view motivated by two lines of empirical evidence. First, there is an overwhelming amount of evidence showing that remembering is often inaccurate, distorted and false. Likely many memories we take to be true, and with which we live our lives seamlessly, are imprecise or wrong. Yet, the evidence also shows that false and distorted memories aren't haphazard, but rather plausible and consistent with a person's acquired experience and the conditions of recall. To illustrate, consider

a classic study by Brewer and Treyens (1981) in which participants were asked to wait in a regular academic office while the researchers worked on setting up the experiment. Unbeknownst to the participants, though, the waiting office itself was the experimental setting. Every object in the office was carefully placed, with some being consistent with the 'schema' of an office (e.g., telephone), and others inconsistent (e.g., cowboy hat). Participants were asked to wait in the office for a little while, and then were taken to a different room for a surprise memory test. In it, participants were given a list of objects, and they were asked to remember which of them were in the office they were just at. The list included both office consistent and office inconsistent 'old' objects—i.e., objects effectively present in the office—as well as 'new' ones, some consistent and some inconsistent with what one would normally find in an academic office. Revealingly, the results showed that participants were more likely to endorse as old (i.e., 'false alarm') new items that were consistent with the schema (i.e., 'lures') relative to inconsistent ones.

Many other influential studies, including well-known manipulations such as the DRM paradigm (Roediger and McDermott 1995)—which produces high false alarm rates to semantically associated word-lures in a study list—as well as the many variations of the 'lost in a shopping mall' study (Loftus and Pickrell 1995)—in which experimenters manage to generate memory experiences of entirely fabricated events that nonetheless are plausible and consistent with the participant's background knowledge and history (Garry et al. 1996)—convincingly demonstrate that false and distorted memories are common and have an air of plausibility or schema-consistency to them.

How can we then square a view of memory as reproductive and of remembering as necessarily linking the encoded content with the retrieved one, with the empirical fact that people often have recollective experiences of items or events they never experienced? Does this mean that our memory system is constantly malfunctioning? Or does it mean that true and false memories are produced by two independent processes? But, if so, why would these two processes be entirely opaque to the subject's awareness? Many researchers argue that the evidence from studies on false and distorted memories speaks against the view that memory is merely reproductive, and suggest instead a view according to which memory should be thought of as reconstructive (Schacter et al. 2000, De Brigard 2014b). According to this view, remembering is not the retrieval of a memory trace where the exact same content stored during encoding is brought back to mind but, instead, it involves the reconstruction of a mental representation aimed at depicting a past event, in a process that may or may not employ stored information acquired in a single past experience.

Now the question is: why would memory be reconstructive? The answer to this question comes, in fact, from the second line of research that has inspired simulationism: the discovery that the neural mechanisms required for episodic memory are also necessary for engaging in certain kinds of imaginations or mental simulations. This line of research dates as far back as 1965, when a classic study on

amnesia by Talland (1965) documented that patients with Korsakoff's amnesia were unable to think about their future. Twenty years later, Tulving (1985) described parallel difficulties in K.C., a famous amnesic case. This observation inspired Tulving to think of episodic memory as a capacity within a larger cognitive system for 'mental time travel', thanks to which we are also able to engage in both episodic past and future thought. Indeed, in the last twenty-five years, the view that our capacity to episodically remember our past and imagine our future are profoundly connected has received substantial support from many scientific fields, including neuropsychology, cognitive neuroscience, and developmental and comparative psychology (Schacter et al. 2015). In particular, it has been consistently shown that mental time travel engages the brain's default mode network (DMN), a set of functionally connected brain regions involving the medial and dorsolateral prefrontal cortices, the posterior cingulate cortex, precuneus and inferior parietal lobule, and the lateral and medial temporal lobes, including the hippocampus (Bucker et al. 2008). More recently, it has also been shown that the DMN is engaged in other kinds of episodic simulations, such as perspective taking (Spreng and Andrews-Hanna 2015) and episodic counterfactual thinking (i.e., our capacity to imagine alternative ways past events could have occurred but did not; De Brigard et al. 2013; De Brigard and Parikh 2019). As a result, some philosophers (De Brigard 2014b; Michaelian 2016) and some neuroscientists (Addis 2020) have argued that the 'DMN is the brain's event simulator' (Addis 2018), and that remembering should be seen simply as a particular operation of this larger episodic simulation system.

Together, these two lines of evidence—one on false and distorted memories and one on the common mechanisms underlying several forms of episodic simulation—motivated some philosophers to reject causalism in favour of a constructivist account in which remembering is a particular instance of a more general capacity to mentally simulate hypothetical personal episodes. Perhaps the most precise articulation of simulationism comes from Michaelian (2024), according to which an individual, *S*, remembers a past event, *e*, if and only if:

1) *S* now represents *e* [*Current representation*] and
2) *S*'s current representation of *e* is produced by a properly functioning and hence reliable episodic construction system that aims to produce a representation of an event belonging to *S*'s personal past [*Proper function*].

Unlike causalism, then, simulationism rejects the need for a past representation condition and for an appropriate causation condition, suggesting instead that all we need is a properly functioning simulation system that can produce reliable representations of past personal events.

Thus characterized, simulationism owes us an explanation as to how exactly should we interpret the proper function condition. As I understand it, Michaelian's interpretation is along the lines of reliabilism in epistemology, so that the episodic

construction system is reliable if and only if it consistently produces true memories (Michaelian 2016). I find this response somewhat unsatisfactory, not only because it makes it a brute fact of the system that it is reliable without explaining how or why, but also because it renders the proper function condition a target of well-known arguments against reliabilism in epistemology (Goldman and Beddor 2021). One such challenge is known as the 'generality problem' (Feldman 1985), and the shape it adopts for the context of memory is that of determining, for any given instance of remembering, which memory forming cognitive process is responsible for its being true. More precisely, the worry is that there is no principled way of individuating the cognitive process of type-remembering such that it could tell us, for a particular token-remembering, whether or not it has been reliably produced.

Lyons (2019) offered a solution to the generality problem in epistemology that is consistent with my own take on the proper function condition for remembering. Essentially, his suggestion is that we can go from type-general to token-specific process if we think of each belief—or, in our case, each memory—as produced by a computational algorithm that is constrained by different parameters. If a particular instance of remembering is generated by a process whereby the values taken up by each variable are within the range of the relevant computational parameters, then that memory has been reliably produced. In other words, the solution for the generality problem is to understand the relevant cognitive process in computational terms. Likewise, in 2014, my own take of the proper function condition in simulationism was to think of the process of memory reconstruction as carried out by a series of computational processes aimed at outputting the optimal solution given their current input.

The devil is in the details, of course, and back then I only glossed over the computational architecture of the reconstructive processes carried by the simulation system (De Brigard 2012, 2014). At that point, the view I endorsed was inspired by the so-called 'rational analysis' of memory (Anderson 1990; for a recent review, see Gershman 2024). According to this approach, cognitive processes can only be understood when considered as adaptations to their environments. In the case of memory, we know for instance that our environment is relentlessly bombarding us with more information than we can perceive, that a lot of the information we manage to perceive won't get stored, and that a lot of the information that gets stored will decay and be forgotten overtime. The data from which we have to reconstruct our memories is thus noisy and incomplete. We also know, though, that there are all sorts of statistical dependencies in our environments: some events are followed by others with high frequency, while others are rather rare or their occurrence is stochastic. The task, then, is to understand how such a computationally limited cognitive system can exploit those statistical dependencies to solve an informational retrieval problem in a way that is adaptive for organisms like us. Moreover, according to the rational analysis, the adaptive purpose of memory is future-oriented, that is, we need to retrieve accurate information about the past for future purposes. Unfortunately,

the process of retrieval is costly. As such, the optimal computational strategy is going to be one that maximizes the odds of a gain (i.e., a successful retrieval of an accurate memory) while minimizing the costs (i.e., failing to retrieve an accurate memory or retrieving an inaccurate one).

This computational framework has the advantage of seeing remembering as a computational process whose purpose is *not* to faithfully preserve and/or reproduce a past content, but to optimally infer from an effect (i.e., the current representation) its most likely cause (i.e., the past representation). In other words, the computational problem our memory system is trying to solve is a variant of what is known as 'an inverse causal problem': the challenge of determining, given a particular effect, what its cause must have been. Given the noisy and incomplete nature of the information the reconstructive process starts off with, the result is going to heavily rely on completions that are highly probabilistic and dependent on background experience and conditions of recall. And now we see how a simulationist account of remembering, whereby memory retrieval is thought as an optimal probabilistic reconstruction of a past experience given both a current noisy content and prior constrains, can accommodate both aforementioned lines of evidence. First the engagement of common neural structures during episodic past, future, and other episodic simulations occurs because they all recruit the same computational constructive processes. Second, the prevalence of schema-consistent false and distorted memories in everyday life is explained by the fact that these very schematic constraints are responsible for the accurate reconstruction of a past experience. Most of the time the mental simulation constructed at retrieval is such that it accurately represents the targeted past event, but sometimes it does not. Nevertheless, in both cases the computational operations underlying the constructive process are identical, and in both cases the system is doing what it is supposed to do.

A Change of Heart

Simulationism can then do away with any need to include a causal claim in their account of remembering, as it does not make it necessary for an accurate memory to include as its content the very same information that the subject experienced in the past and is now remembering. A genuine memory could just as well be produced by the same computational processes without the need to include information directly caused by the original event. Moreover, simulationism can also remove the need to postulate memory traces, understood as preserved stand-ins for the encoded content, poised to be recovered at the time of retrieval. If an accurate memory is fully reconstructed at retrieval, talk about memory traces may become unnecessary.

However, in the decade since I published my own version of simulationism (De Brigard 2014b), it has become evident that there are a number of empirical and conceptual issues that put pressure against some core claims I defended back

then—although it is likely that these criticisms affect other versions of simulationism as well, such as Michaelian's (2016). What I seek to do in this section is to offer both empirical and conceptual reasons against two such core claims, namely that the brain's DMN is an episodic construction system (it isn't), and that we don't need memory traces to explain remembering (we do).

Simulationism 2.0

According to simulationism, remembering is produced by a single cognitive system whose function is to generate episodic simulations, with memories being but a subset of them. Such an episodic construction system, the view goes, corresponds to the brain's DMN. After all, thinking of the DMN as the neural structure subserving episodic simulation helps to explain the commonalities between episodic recollection and other kinds of episodic simulations that emerge in parallel during development and are equally affected in individuals with brain damage. For instance, individuals with hippocampal damage have difficulty generating episodic memories and also episodic future and counterfactual thoughts (Schacter et al. 2015). Likewise, individuals with damage in the medial prefrontal cortex have difficulty spontaneously retrieving episodic autobiographical memories (Belfi et al. 2018) as well as generating episodic hypothetical thoughts (Beldarrain et al. 2005). Indeed, the idea that self-projection or self-simulation constituted a unified psychological kind, became an attractive theory to explain the function of the DMN when it was initially functionally characterized (Carroll and Buckner 2007).

Unfortunately, there is quite a bit of counterevidence that is hard to accommodate within this view. First of all, there are now many experimental results showing that activity in the DMN is associated with all sorts of cognitive functions that are hard to fit under the umbrella of 'episodic simulation', including semantic processing (Lanzoni et al. 2020), allostasis and interception (Kleckner et al. 2017), addiction (Zhang and Volkow 2019), and aesthetic appeal (Belfi et al. 2019), among others. It would be a stretch, I think, to try to group all of these cognitive functions under the same general category of 'episodic simulation'. Alternatively, if everything counts as 'episodic simulation', then the explanatory advantage of the construction system in explaining remembering would be severely diminished.

A second concern is that many non-human animals have a DMN, yet it is very questionable whether they can engage in complex episodic simulation. Neural evidence has revealed the presence of DMN in the brain of mice (Sforazzini et al. 2014), rats (Lu et al. 2012), rabbits (Schroeder et al. 2016), marmosets (Liu et al. 2019), and monkeys (Vincent et al. 2007). The presence of this brain network across such vastly different mammalian species suggests that it may not have evolved for complex cognitive functions such as the generation of episodic counterfactual thoughts or autobiographical recollections. A more parsimonious explanation is perhaps

that the DMN reflects more basic metabolic and homeostatic processes, as initially hypothesized by Raichle and colleagues (2001), rather than constituting a unified cognitive system whose function is best characterized in psychological terms.

A third issue with the claim that the DMN is the simulation system of the brain is the fact that not all episodic constructive processes depend to the same extent of core regions of the DMN—especially the hippocampus. As it turns out, the engagement of the DMN in the construction of episodic simulations is highly dependent on the particular contents that are simulated. Research on episodic counterfactual thinking, for instance, has shown that thinking about alternative ways in which events could have unfolded in one's past recruits the DMN, but not so when it comes to thinking about what an unfamiliar person in an unfamiliar situation would have done (De Brigard et al. 2015). Likewise, the DMN is *not* preferentially recruited when it comes to thinking about episodic counterfactual situations that involve objects, as opposed to people (Parikh et al. 2018). Similar considerations apply to episodic simulations involving theory of mind, as patients with hippocampal damage can engage in such simulations as long as they do not demand retrieving information from episodic memory (Rosenbaum et al. 2007).

A fourth issue concerns navigation. A critical function of the DMN and, in particular, the hippocampus, is the capacity to mentally generate spatial simulations (Spreng et al. 2009). But even this association, I surmise, is questionable. For years, it's been thought that medial temporal lobe structures—mainly the hippocampus and entorhinal cortex—are both necessary and sufficient for navigation. However, in a recent study, Long and Zhang (2021) demonstrated spatial mappings in the somatosensory cortex of foraging rats, and shortly after Wikenheiser and colleagues (2021) found spatially localized firing in neurons in the rat's orbitofrontal cortex that mimic those found in the hippocampus—the only difference being that they were less temporally precise. Critically, both the somatosensory and the orbitofrontal cortices are not part of the canonical DMN. Moreover, evidence from individuals with developmental as well as adult-onset amnesia reveals several preserved spatial navigation abilities despite their episodic memories being impaired (Rosenbaum et al. 2000; Rosenbaum et al. 2015). In fact, a very recent study of a patient with severe bilateral medial temporal lobe damage shows comparable performance on spatial navigation and spatial memory tasks despite abysmal results in autobiographical and episodic memory tests (McAvan et al. 2022).

Fifth, and perhaps more relevant for our current purposes, is the fact that individuals with hippocampal damage can still generate spatial simulations. One of the leading theories seeking to explain the common engagement of the hippocampus during episodic past, future, and other hypothetical simulations, has been defended by Maguire and colleagues (e.g., Maguire and Mullaly 2013). According to this view, the hippocampus is required for us to be able to engage in the mental simulation of scenes in space. The problem is that even their own data suggests that patients with hippocampal damage can still think in and about space. In their landmark paper,

Hassabis et al. (2007) showed that patients with hippocampal damage had difficulty thinking about new experiences, and they interpret such deficits as reflecting their incapacity to mentally simulate spatial scenes. But take a look at the transcript of one of the (available) narratives from one of the patients, who is asked to imagine standing in the main hall of a museum:

> *Interviewer*: So, what does it look like in your imagined scene?
> *Patient P05*: Well, there is big doors. The openings would be high, so the doors would be very big with brass handles, the ceiling would be made of glass, so there's plenty of light coming through. Huge room, exit on either side of the room, there is a pathway and map through the centre, and on either side there'd be the exhibits.

How can one read this transcription and not think that this patient is mentally simulating a scene that occurs in space? How can one imagine that some things are on one side of a room, or that light comes through the ceiling, or that there is a map in the centre of the room, if one can't envision spatial scenes? It is certainly the case that, when compared with controls, these mental simulations are less rich and detailed—a point I'll discuss below—but it would be false to say that they aren't reflecting an imagined spatial scene. Moreover, there is evidence that even H.M. was able to not only remember spaces from his old house, but he was also capable of mentally navigating through such spaces (Corkin 2013).

And finally, contrary to some interpretations of the mental time travel system, it turns out there is also evidence to the effect that the hippocampus may not be required for thinking about time. Recent studies on K.C., the famous amnesic patient that motivated Tulving to talk about mental time travel in the first place, have shown that he has no problem talking about temporal concepts such as time travel, why events that happened in the past can't be modified in the present, why the future does not affect the past, or why people wouldn't do things now because they may regret them in the future (Craver et al. 2014a). K.C. is also subject to decision-making biases that allegedly require the capacity to anticipate the future, such as the Allais paradox (Craver et al. 2014b)—the tendency to give inconsistent answers in violation of expected utility theory when forced to choose between two gambits—and delayed discounting in intertemporal choice (Kwan et al. 2015). Neuroimaging evidence supports these findings, showing neural differences in brain activity associated with episodic future thinking versus delayed discounting in economic choices (Benoit et al. 2011).

Taken together, these various findings strongly suggest that the DMN is *not* a distinct cognitive system for the mental construction of episodic simulations. On the contrary, the evidence indicates that it is neither necessary—for it is possible to generate the kinds of contents such a system is supposed to produce without the engagement of core regions of the DMN—nor sufficient for episodic simulation, as some kinds of mental constructions of imaginative scenes recruit regions outside

of the DMN.[1] In my view, what the available evidence supports is a much narrower claim: namely, than a healthy hippocampus—and hippocampal path (e.g., Ayala et al. 2022)—is needed to successfully engage in episodic recollection and certain kinds of episodic simulations. The question now is: what do these episodic simulations have in common?

A possible answer comes from thinking about the common engagement of the hippocampus during these kinds of episodic simulations as having less to do with *what* they represent—i.e., what their contents are about—and more with *how* they represent it, i.e., the representational format or structure. The suggestion I put forth (and which I more fully articulated in De Brigard and Gessell 2016), is that the kinds of episodic simulations relevant for understanding the neural substrates of mental time travel can vary along two dimensions: content and structure. Now, contents can be more or less about time; some of our thoughts are clearly about past or future events, some are clearly not, and some are in between. Call this dimension *tense*. But the structure of our thoughts can also vary along a temporal dimension: some thoughts take little time to be entertained and they don't seem to require much of a temporal structure to unfold; others, by contrast, need more time as their contents involve structures that need time to unravel. Call this dimension *dynamicity*. Remembering that Caracas is the capital of Venezuela, or what an apple looks like, are relatively static in that they don't require much time unfolding in working memory to be entertained, whereas remembering one's first kiss, imagining how things may have unfolded had one not missed a particular elevator ride, or thinking about fixing the dishwasher in the weekend, presumably require more. The more dynamic the structure of the mental simulation, I surmise, the more the hippocampus is needed.

As reviewed above, there is evidence suggesting that core regions of the DMN—and, in particular, the hippocampus—are neither necessary nor sufficient to engage in thoughts *about* space or time. But if we think of the hippocampus—at least in the human case—as required for episodic simulations that unfold *over* time, then the counterevidence is no longer threatening. Most of the tensed thoughts K.C. is capable of tend to be short and somewhat telegraphic, not unlike the reports from the amnesic patients in the Hassabis et al. (2007) study. For instance, patient P03, when asked to imagine 'lying on a white sandy beach in a beautiful tropical bay', simply responds 'all I can see is the colour of the blue sky and the white sand'. Further probing elicits no more than that 'like I'm kind of floating'. Similar to the transcript of patient P05 above, what these two narratives seem to have in common is that they lack dynamicity; it is as if they were describing a static picture in space, rather than immersing themselves in an episodic simulation that's unfolding over time in

[1] A reviewer suggested that there is evidence to the effect that the DMN actually consists of at least three sub-networks (Andrews-Hanna et al. 2010), and that it may be possible that while the DMN as a whole isn't necessary or sufficient for episodic simulation, perhaps one such sub-region may be. While it may be possible that some parcellation of the DMN has a better chance at being identified as 'the simulation system of the brain' (although I am very sceptical), my target here is the original formulation of simulationism, which was not confined to a sub-network of the DMN.

their minds. Indeed, additional evidence shows that individuals with hippocampal damage can describe spatial scenes, as long as they are static (Gaesser et al. 2011; Race et al. 2011; Race et al. 2013), as well as non-tensed fictional events that do not require much dynamicity or elaboration (Rosenbaum et al. 2009; Romero and Moscovitch 2012).

This dynamic interpretation of the role of the hippocampus in episodic simulation also helps to account for some classic neuroimaging results in the mental time travel literature. For instance, in their landmark paper, Szpunar et al. (2007) found no difference in hippocampal engagement between their experimental conditions (i.e., episodic past and future thinking) and their control condition, whereas Addis et al. (2007) did. The reason, I surmise, is because in the former, the control condition was an episodic simulation involving a familiar other (i.e., Bill Clinton), whereas in the latter, the control condition was a sentence construction task, which does not require the dynamic deployment of a complex mental simulation. The recruitment of the hippocampus during non-mental time travel tasks, such as imagining fictitious (Hassabis et al. 2007) and non-temporal events (D'Argembeau et al. 2008), as well as possible personal past (Addis et al. 2009) and counterfactual episodes (De Brigard et al. 2013; van Hoeck et al. 2013), can also be explained by the fact that such simulations involve the generation of complex dynamic representations. And, incidentally, thinking about the role of the hippocampus in the generation of dynamic, as opposed to static, simulations, may help to resolve the theoretical conflict between the *scene construction* (Hassabis and McGuire 2007) and the *constructive episodic simulation* (Schacter and Addis 2007) hypotheses. Scene construction is defined as 'the process of mentally generating and maintaining a complex and coherent scene or event' (Hassabis and McGuire 2007: 299). The view of the hippocampal role in episodic simulation I advocate for here is entirely compatible with the scene construction perspective insofar as the scene that is mentally generated and maintained is dynamic, structured and unfolds over time. This makes the scene construction view a hypothesis about the structure or the format of the mental simulation. By contrast, the episodic simulation hypothesis speaks to the etiology of the contents of our episodic simulation—whether they are drawn from episodic memories, or semantic and conceptual knowledge—not about their format. Both views are, then, compatible.

Finally, it is worth noting that although the hippocampus is traditionally associated with spatial representations, thanks to the discovery of place and grid cells, there is also quite a bit of evidence showing that these cells are also sensitive to sequential information about spatial navigation. Landmark studies by O'Keefe and Recce (1993) and Skaggs and colleagues (1996) showed that the moment of the firing of a place cell within a navigational sequence has a precise timing relation with oscillations in the theta band. Likewise, Foster and Wilson (2007) demonstrated that place cells in CA1 are 'timed-locked' to theta oscillations, suggesting that, prior to performing a learned sequence, place cells can 'pre-play' the upcoming action by firing in succession. These results suggest that the hippocampus, and the entorhinal

cortex as well, are not only encoding the spatial but also the temporal relations among the components of a scene. The fact that the components of a simulated episode are bounded by temporal relations whose mental reinstatement requires time to unfurl may be as essential to the engagement of the hippocampus as the fact that the elements of the simulation stand in spatial relations too.

In sum, the purpose of this section has been twofold. First, I offered some counterevidence against the simulationist's claim that the DMN is the 'episodic simulation system of the brain' (De Brigard 2014b; Addis 2018, 2020). I argued, instead, that although the evidence does not point towards there being a single unified system for episodic simulation, it does suggest that the hippocampus—and likely adjacent regions in the medial temporal lobe—are required for the proper construction and maintenance of certain kinds of episodic mental simulations. What these kinds of episodic mental simulations have in common—and this is the second point I sought to put forth in this section—is that they are dynamic, in the sense that their contents are not only spatially but also temporally structured in a way that the very unfolding of the episodic simulation (i.e., its generation and maintenance) is protracted and takes time. Thus, simulationism needs to abandon the idea that there is a unified cognitive system for episodic simulation and, instead, adopt the narrower view that the commonality in hippocampal engagement has to do with the dynamic format of the episodic simulation (De Brigard and Gessell 2016).

Probabilistic Dispositionalism

Unlike causalism, simulationism sees no need for the appropriate causation condition. After all, the episodic construction system can output a reliable memory that neither represents nor was caused by an actual past event (Michaelian 2024). This also means that simulationism has no use for memory traces. Recall that the notion of memory trace was introduced as a content-bearing theoretical entity that could help to bridge the causal gap between the encoded (i.e., past representation) and the remembered events (i.e., current representation). But if there is no prior representation, then there is no gap to be bridged, and thus there is no need for memory traces (Michaelian 2016). Memory traces, therefore, appear to be incompatible with simulationism.

In this section, however, I argue against this claim. My argument has a negative and a positive part. In the negative part, I argue that, contrary to what simulationism holds, there are many memory-related phenomena whose explanation makes it indispensable to appeal to memory traces. In the positive part, though, I argue that there is a promising account of the nature of memory traces that not only makes sense of why the hippocampus is involved in several kinds of episodic simulations (including remembering) but also why is it that false and distorted memories can be the normal product of a properly functioning reconstructive process. For reasons

that I hope will become clearer in what follows, I call this view on memory traces 'probabilistic dispositionalism'.

There are at least two arguments for thinking that memory traces are indispensable when it comes to explaining the psychological process of remembering. The first one, which I've developed at length before (De Brigard 2020), has to do with differential effects in recollection and it is directed against a well-known argument against memory traces from Malcolm (1977). Many philosophers—including Malcolm—agree that memory traces are postulated via inference to the best explanation in order to avoid either the metaphysically dubious direct realist route or the equally problematic acceptance of causation-at-a-temporal distance between a non-existent past event (i.e., the rember*ed* event) and a current one (i.e., the remember*ing* event; Heil 1978; Bernecker 2008). Since the justification for their postulation is an inference to the best explanation, then it follows—according to Malcolm—that accounts of remembering that make use of the notion of memory trace should be better than those that do not. But then Malcolm suggests that when I give an account of why is it that I remember a particular event, say, a boat capsizing, all I need to do is refer to my prior experience of having witnessed the boat capsizing. At no point do I need to invoke, in addition to my having witnessed the event, the existence of some unobservable causal intermediary memory trace. To not multiply entities without necessity, and given that both accounts are equally good, then there is no reason to postulate the existence of memory traces.

The problem with this argument is that, in only looking at cases of successful recollection, Malcolm overlooks a large swathe of memory related phenomena for which talk about intermediary memory traces becomes indispensable. For in addition to successful recollection, we often demand causal explanations for cases of *unsuccessful* recollection. Suppose that I witness a boat capsizing alongside my friend Andrea. The following day, I recall the event but Andrea doesn't. It seems as though some causal story is needed in order to explain why is it that I can remember the event whereas Andrea cannot. Or suppose that I can remember more details than she can, or that my memory is profoundly distorted relative to hers. In a word, cases of differential recollection under the same encoding conditions highlight the need to include some story about causal intermediaries that can explain the difference at retrieval. Appealing to memory traces—or, if you are unhappy with the term, some kind of causal intermediary between encoding and retrieval—becomes indispensable again.

The second argument is related, though it involves the use of a pharmacological agent—propranolol—during memory reactivation. One of the most interesting findings in the memory literature in the past few decades, is the fact that when a memory is reactivated at retrieval, it becomes labile and prone to modification (Nader and Einarsson 2010). The evidence suggests, in fact, that there is a window of time in which, if the reactivated memory is intervened upon, its contents can be modified, even erased. A pharmacological intervention using propranolol, a synthetic beta-adrenergic receptor blocker that acts as an inhibitor of protein synthesis

underlying memory consolidation and re-consolidation, has been shown to successfully extinguish both stimulus and context specific conditioning when administered immediately after memory reactivation and prior to reconsolidation (e.g., Leal Santos et al. 2021). Importantly, these effects are condition-specific, meaning that the administration of propranolol only affects the retention of memories tied to a particular spatiotemporal context. Other memories that may have been acquired before or after, are unaffected. Given the specificity of these effects, it seems extremely difficult to explain such results without assuming that something in the brain of the animal was altered, and that such neural substrate is tied to a specific experienced event. Once again, the need to posit memory traces—or at least intermediate causal mechanisms—becomes evident.

The explanatory indispensability of a theoretical term is certainly not a sufficient condition to accept the existence of its putative referent. Nevertheless, in the case of memory traces, it has been a good motivator to try to discover them, and one that clearly has inspired neuroscientists for decades to understand what their nature might be (Josselyn and Tonegawa 2020). My purpose, in the rest of this section, is to offer a general framework to characterize the nature of memory traces such that it can accommodate both the fact that memories are often distorted and false and the narrower thesis, defended in the previous section, that the hippocampus is required for the successful construction of dynamic episodic simulations, including episodic memories.

But first: some conceptual clarifications. Views on memory traces vary along several dimensions. On the one hand, some of those dimensions affect the *vehicle* of the mnemonic representation. For some researchers, memory traces are localized representations, akin to discrete symbolic entities carrying a particular mnemonic content, and instantiated in dedicated neuronal populations or even in specific cells (Gallistel 2010). For others, memory traces are distributed across neuronal connections (Rumelhart and McClelland 1986), and for some they are a combination of both (Gershman 2023). Moreover, some even think of memory traces as extended and/or embodied, meaning that the vehicles carrying the relevant mnemonic content extend beyond the limits of the brain. On the other hand, some dimensions pertain to the degree of explicitness of the mnemonic content. Some hold that memory traces carry stripped-down versions of the encoded content but they do so explicitly, meaning that, in principle, one could read off the content directly from the neuronal substrate. For other, contents are implicitly encoded, for an additional process—i.e., retrieval—is needed to make them explicit. And, finally, some argue that contents are not occurrent—and, thus, are neither explicitly nor implicitly encoded—but rather dispositional, i.e., what gets encoded is a disposition to revive a content at retrieval given the right cue. Each one of these views has advantages and disadvantages, the discussion of which unfortunately I have to sidestep (but see De Brigard 2014a and De Brigard 2023). Nevertheless, I hope this brief excursus on the terminology is sufficiently informative, as the view I seek to defend in what follows is a version of dispositionalism predicated on representational vehicles, not on

contents, whereby a memory trace is the disposition of a neural network to reactivate, as approximately as possible, the state it was in during encoding at the time of retrieval. Let's see how this works.

The framework I have in mind builds upon the hippocampal indexing theory (HIT), initially put forth by Teyler and DiScena (1986), in order to explain how memory traces formed during encoding could be reinstated at retrieval. Consistent with the complementary learning systems model (McClelland et al. 1995), HIT postulates that when an event is experienced, two consolidation processes take place. First, there is a fast *cellular* consolidation in which information is encoded as changes in connectivity among the neurons involved in the initial processing of the encoded event. Second, there is the more protracted *systems* consolidation, in which hippocampal-neocortical connections are further strengthened (Figure 9.1).[2] As an illustration, consider how the model explains the formation of a memory trace for, say, the experience of seeing a boat capsizing. Walking by the shore you peer at the horizon and notice a vessel tipping over. This experience engages several regions of your cortex: sounds will be processed by the auditory cortex, shapes, colours, and the like will be processed by the visual cortex, smells by the olfactory cortex, and so forth. Active neurons in the hippocampus form an *index* (more about this term in a second) that binds these distributed cortical patterns into a larger hippocampal-neocortical network which, over time, becomes systems-consolidated. Now suppose that a few days later, someone asks you if you've ever seen a boat capsizing. This auditory cue helps to re-activate a portion of the cortical pattern, and this reactivation propagates to the hippocampal index, which in turn enables the reactivation of the rest of the neuronal pattern, effectively reinstating the encoded hippocampal-neocortical network. Since the vehicle of the encoded representation is reactivated, then the encoded content is reenacted and, thus, you manage to remember the event.

The notion of *index* requires further clarification. As far as I know, the first reference to a hippocampal index is from Marr (1971), who called it a 'simple memory', meaning a kind of sketchy or abstract representation of the encoded event. According to this view, what the hippocampus does is store a low-dimensional representation—perhaps akin to a lossy compression format (e.g., JPEG for images) or a set of principal components—out of which the encoded representation can be reconstructed. I tend to disagree with this view, for three reasons. The first reason is that compression formats require decompressing processes, and it is unclear what would be doing the decompressing in the hippocampal case. The second reason is that having a second representation that allegedly encodes more or less the same information as the first one, makes the hippocampal activity rather redundant.

[2] When initially proposed, HIT followed the standard model of memory according to which, once consolidated, the hippocampus was no longer required for retrieval. However, further examination of extant data as well as new evidence indicates that the hippocampus is still required for the retrieval of recent and remote memories (Nadel and Moscovitch 1997). The version of HIT I advocate for here is consistent with this further development and has been further incorporated into a recent update of the CLS view (Kumaran, Hassabis, and McClelland 2016).

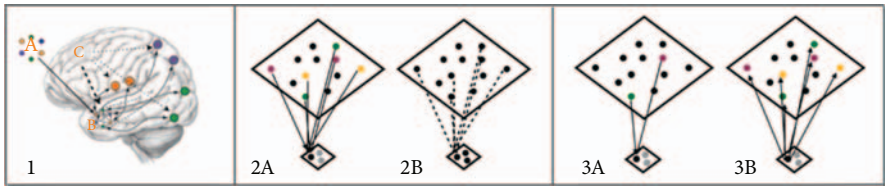


Figure 9.1 Hippocampal Indexing Theory [HIT]. **(1) Graphical schematic in brain space.** (A) An initial stimulus with multiple sensory properties is first experienced. (B) A rapid consolidation occurs in the hippocampus while the sensory information of the stimulus is processed in the relevant areas of the sensory cortex (i.e., visual cortex for visual properties, auditory cortex for auditory properties, etc.). This co-activation creates an association between the sensory regions and a hippocampal index. (C) At retrieval, a top-down signal from the pre-frontal cortex to elements of the hippocampal-cortical assembly reactivates the network and, thus, the memory content. **(2) Encoding**. (2A) The bigger layer indicates units in the neocortex, with different colors indicating different sensory information. The smaller layer indicates specific synaptic activity uniquely associated with the pattern of neocortical activation. (2B) After encoding, consolidation strengthens the connection between the hippocampal index and the associated neurons in the neocortex. **(3) Retrieval**. (3A) A cue can reactivate a subset of the neocortical pattern, which in turn reactivates the hippocampal index. (3B) This reactivation further spreads to the rest of the hippocampal-neocortical network, effectively reinstating the encoded pattern. (Original figure from De Brigard 2023).

And, finally, it seems unlikely that all the neocortical variability that is required to represent different modalities and formats of information can be captured by the sparser archicortex of the much smaller structure that the hippocampus is. In fact, when trying to decode categorical and sensory information from brain activity at retrieval, multivariate patterns are unable to recover any encoded structure from hippocampal activity alone (Huffman and Stark 2014). The evidence, therefore, does not suggest that the hippocampus is in fact storing a 'simple memory'.

Instead, I suggest that what the hippocampal index encodes is not an explicit—if compressed—content but rather a set of conditional instructions or dispositions to reactivate, as best as it can, the cortical pattern of activity it was associated with during encoding. Thus, when one experiences a certain event during encoding, the experienced content is instantiated in a particular representational vehicle, in the form of a hippocampal-neocortical network in the brain. Consolidation increases the probability of the nodes in the network to coactivate given the right cue. When such a cue is presented in the retrieval context, the coactivation among units of the network starts to propagate towards the hippocampal index, which does not contain explicit contents but rather the conditional instructions to reactivate the rest of the pattern of activity[3]. This, I suggest, is the right way to understand what a

[3] One can think about the end state of the reactivation in terms of a Hopfield network (Hopfield 1982), a kind of recurrent artificial network that tend to stabilize in a particular pattern of activation or 'attractor'. The idea, then, is that the final state of reactivation of the hippocampal-neocortical network is the attractor state of the network that underlies the encoded content.

memory trace is: the dispositional property of a neural network to reinstate the state it was in, during encoding, at the time of retrieval. In fact, by characterizing the memory trace as a dispositional property of a representation vehicle rather than a content, one not only avoids concerns about content dispositionalism (Vosgerau 2010), but also can readily explain why unexpected cues can bring to mind involuntary memories (Berntsen 2009). Additionally, this view helps to explain why, when a memory is reactivated, it becomes modifiable. Since every act of retrieval is itself an act of re-encoding, nodes that weren't part of the original pattern but that already have a higher baseline probability of being coactive, are now more likely to getting included in the original pattern of activation after reconsolidation. Finally, this view also accommodates the fact that information acquired after the encoding and prior to retrieval can influence the way we remember past events. Here's an example I like to use. Long ago, before I learned how to speak English, I committed to memory the chorus of 'A hard days' night' by the Beatles. I did not know what it meant, but I could sing the words. Years later, after learning English, I found myself listening to the song again, and was able to remember the lyrics and sing along. But as I was remembering the words, I was also understanding them for the very first time. The content of my recollection was different from that of its encoding, due to an intervening change to the network units that formed the representational vehicle of my memory.

Rapprochement

In the previous section I sought to defend two claims. First, contra simulationism, I argued that memory traces are often explanatorily indispensable to account for certain instances of remembering and memory related phenomena. Second, I offered a general framework, inspired by the HIT, as to how to understand memory traces and the role the hippocampus plays at retrieval. Specifically, I suggested that memory traces could be understood as dispositional properties of hippocampal-neocortical networks to reinstate the pattern of activation they were in, during encoding, at the time of retrieval. Since this pattern of activation consists in a set of nodes whose probability of coactivation is high—perhaps relative to some threshold—the reactivation is going to be sometimes imprecise and noisy.

Thinking about memory traces in this way—call this approach 'probabilistic dispositionalism'[4]—has the advantage of accommodating the two main motivations for simulationism. First, the fact that false and distorted memories are schema-consistent is a natural consequence of the probabilistic pattern completion process. As mentioned, a lot of the information that is initially perceived won't be encoded, and retention may in turn degrade some of the connection weights between units. But thanks to the statistical regularities in the connections of such units, the

[4] For similar views, see Vosgerau 2010; Perrin, 2018; Perrin 2021; Werning 2020.

probability of reactivating the right set of connections given a cue remains high. This is the sense in which reconstruction can be said to be both probabilistic and veridical. However, since units in the hippocampal-neocortical network have additional existing associations with other units, it is possible that a unit or a set of units not involved in the encoding of the original content can become active during pattern completion at retrieval, resulting in a reconstructed memory that does not accurately represent the past event.[5]

Second, probabilistic dispositionalism can also explain why the hippocampus is engaged in certain kinds of episodic simulation, besides remembering. If we think of the hippocampal index as triggering the reactivation of different sensory areas in order to reenact the contents they process, and such reactivation takes time to complete into a single coherent scene, then we can think of this computation in analogous terms for episodic memory as well as for other dynamic mental simulation. More precisely, since the hippocampal index per se does not represent the information processed in the neocortex but rather rules to reactivate a more complex set of cortical units responsible for actually carrying out the contents of the mental simulations, the same reactivating computations can be recruited to recreate dynamic scenes that do not correspond to actual past but to hypothetical episodes. As such, it is not surprising that similar hippocampal-neocortical networks are recruited during certain kinds of episodic simulations because the computations that underlie their generation are not that dissimilar from those involved in episodic recollection.

Although there is much more that can be said about the computational nature of memory traces in general, and about my favoured probabilistic-dispositional account in particular, I hope that what I've said enable us to see how we may be able to dissolve the conflict between causalism and simulationism. For simulationism need not postulate the existence of a dedicated system for episodic simulation, but simply accept that certain structures—particularly the hippocampus—are shared among certain kinds of episodic simulations. Such commonalities are explained because the same underlying computations for dynamic pattern completion can be redeployed in certain kinds of episodic simulation. Additionally, simulationism can accept the existence of memory traces, as long as they are understood as dispositional properties of hippocampal-neocortical networks (i.e., representational vehicles) to probabilistically reenact the state they were in, during encoding, at the time of retrieval. Now, whether probabilistic dispositionalism can help to understand our tendency to think of remembering in causal terms is, alas, a story I need to leave for another day.[6]

[5] There is quite a bit of research in computational psychology and neuroscience trying to understand the precise computations that best characterize the probabilistic process of pattern completion at retrieval. In my opinion, a promising avenue is the 'rational analysis approach' (Anderson 1990) I mentioned before. On this perspective, extant associations between units can influence the pattern of neural activation by combining values reflecting prior frequencies as well as previously acquired conceptual/semantic associations.

[6] Previous versions of this paper were presented at the conference on *Simulationism* at the University of Grenoble, Alpes, in July 2022, the workshop on *Memory, Space, and Time* at the University of Arizona, in November 2022, and the *Generative Episodic Memory* conference at Bochum, in June 2023. Many thanks to the organizers and the audiences of these events.

References

Addis, D. R., Wong, A. T., & Schacter, D. L. (2007). Remembering the past and imagining the future: Common and distinct neural substrates during event construction and elaboration. *Neuropsychologia*, 45(7), 1363–77.

Addis, D. R., Pan, L., Vu, M. A., Laiser, N., & Schacter, D. L. (2009). Constructive episodic simulation of the future and the past: Distinct subsystems of a core brain network mediate imagining and remembering. *Neuropsychologia*, 47, 2222–38.

Addis, D. R. (2018). Are episodic memories special? On the sameness of remembered and imagined event simulation. *Journal of the Royal Society of New Zealand*, 48, 64–88.

Addis, D. R. (2020). Mental time travel? A neurocognitive model of event simulation. *Review of Philosophy and Psychology* (Special Issue on Mental Time Travel), 11, 233–59.

Anderson, J. R. (1990). *The Adaptive Character of Thought*. Hillsdale, NJ: Psychology Press.

Andrews-Hanna, J. R., Reidler, J. S., Sepulcre, J., Poulin, R., & Buckner, R. L. (2010). Functional-anatomic fractionation of the brain's default network. *Neuron*, 65(4), 550–62.

Ayala, O. D., Banta, D., Hovhannisyan, M., Duarte, L., Lozano, A., García, J. R., … & De Brigard, F. (2022). Episodic past, future, and counterfactual thinking in relapsing-remitting multiple sclerosis. *NeuroImage: Clinical*, 34, 103033.

Bartlett, F. C. (1932). *Remembering: A Study in Experimental and Social Psychology*. Cambridge, UK: Cambridge University Press.

Beldarrain, M. G., Garcia-Monco, J. C., Astigarraga, E., Gonzalez, A., & Grafman, J. (2005). Only spontaneous counterfactual thinking is impaired in patients with prefrontal cortex lesions. *Cognitive Brain Research*, 24(3), 723–26.

Belfi, A. M., Karlan, B., & Tranel, D. (2018). Damage to the medial prefrontal cortex impairs music-evoked autobiographical memories. *Psychomusicology: Music, Mind, and Brain*, 28(4), 201.

Belfi, A. M., Vessel, E. A., Brielmann, A., Isik, A. I., Chatterjee, A., Leder, H., … & Starr, G. G. (2019). Dynamics of aesthetic experience are reflected in the default-mode network. *NeuroImage*, 188, 584–97.

Benoit, R. G., Gilbert, S. J., & Burgess, P. W. (2011). A neural mechanism mediating the impact of episodic prospection on farsighted decisions. *Journal of Neuroscience*, 31(18), 6771–79.

Bernecker, S. (2008). *The Metaphysics of Memory* (Vol. 111). UK: Springer Science & Business Media.

Bernecker, S. (2009). *Memory: A philosophical study*. Oxford: Oxford University Press.

Berntsen, D. (2009). *Involuntary Autobiographical Memories*. New York, USA: Cambridge University Press.

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive psychology*, 13(2), 207–30.

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Annals of the new York Academy of Sciences*, 1124(1), 1–38.

Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in Cognitive Sciences*, 11(2), 49–57.

Corkin, S. (2013). *Permanent Present Tense: The Unforgettable Life of the Amnesic Patient H. M*. NY: Basic Books.

Craver, C. F., Kwan, D., Steindam, C., & Rosenbaum, R. S. (2014a). Individuals with episodic amnesia are not stuck in time. *Neuropsychologia*, 57, 191–95.

Craver, C. F., Cova, F., Green, L., Myerson, J., Rosenbaum, R. S., Kwan, D., & Bourgeois-Gironde, S. (2014b). An Allais paradox without mental time travel. *Hippocampus*, 24(11), 1375–80.

D'Argembeau, A., Xue, G., Lu, Z. L., Van der Linden, M., & Bechara, A. (2008). Neural correlates of envisioning emotional events in the near and far future. *Neuroimage*, 40(1), 398–407.

De Brigard, F., & Gessell, B. S. (2016). Time is not of the essence: Understanding the neural correlates of mental time travel. In Michaelian, K., Klein, S. B., and Szpunar, K.K. (eds). *Seeing the Future: Theoretical Perspectives on Future-Oriented Mental Time Travel* (pp. 153–179). Oxford.

De Brigard, F., & Parikh, N. (2019). Episodic Counterfactual Thinking. *Current Directions in Psychological Science*. 28(1), 59–66.

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), 2401–14.

De Brigard, F., & Giovanello, K. S. (2012). Influence of outcome valence in the subjective experience of episodic past, future, and counterfactual thinking. *Consciousness and Cognition*, 21(3), 1085–96.

De Brigard, F., Spreng, R. N., Mitchell, J. P., & Schacter, D. L. (2015). Neural activity associated with self, other, and object-based counterfactual thinking. *Neuroimage*, 109, 12–26.

De Brigard, F. (2014a). The nature of memory traces. *Philosophy Compass*, 9(6), 402–14.

De Brigard, F. (2014b). Is memory for remembering? Recollection as a form of episodic hypothetical thinking. *Synthese*, 191(2), 155–85.

De Brigard, F. (2019). Know-how, intellectualism, and memory systems. *Philosophical Psychology*, 32(5), 720–59.

De Brigard, F. (2020). The explanatory indispensability of memory traces. *The Harvard Review of Philosophy*, 27.

De Brigard, F. (2023). *Memory and Remembering*. Cambridge University Press.

Feldman, R. (1985). Reliability and justification. *The Monist*, 68(2), 159–74.

Foster, D. J., & Wilson, M. A. (2007). Hippocampal theta sequences. *Hippocampus*, 17(11), 1093–99.

Furlong, E. J. (1948). Memory. *Mind*, 57, 16–44.

Gaesser, B., Sacchetti, D. C., Addis, D. R., & Schacter, D. L. (2011). Characterizing age-related changes in remembering the past and imagining the future. *Psychology and Aging*, 26(1), 80.

Gallistel, C. R., & King, A. P. (2010). *Memory and the Computational Brain*. Wiley-Blackwell.

Garry, M., Manning, C. G., Loftus, E. F., & Sherman, S. J. (1996). Imagination inflation: Imagining a childhood event inflates confidence that it occurred. *PBR*, 3, 208–14.

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

Gershman, S.J. (2024). The rational analysis of memory. In M. Kahana & A. Wagner (eds.), *Oxford Handbook of Human Memory*. Oxford University Press.

Goldman, A., & Beddor, B. (2021). Reliabilist Epistemology. *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/sum2021/entries/reliabilism/.

Hassabis, D., & Maguire, E. A. (2007). Deconstructing episodic memory with construction. *Trends in Cognitive Sciences*, 11(7), 299–306.

Hassabis, D., Kumaran, D., Vann, S. D., & Maguire, E. A. (2007). Patients with hippocampal amnesia cannot imagine new experiences. *PNAS*, 104, 1726–31.

Heil, J. (1978). Traces of things past. *Philosophy of Science*, 45(1), 60–72.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8), 2554–58.

Huffman, D. J., & Stark, C. E. (2014). Multivariate pattern analysis of the human medial temporal lobe revealed representationally categorical cortex and representationally agnostic hippocampus. *Hippocampus*, 24(11), 1394–403.

Hume, D. (1748). *A Treatise of Human Nature.* Oxford.

Hume, D. (1748/1777). *An Enquiry Concerning the Human Understanding*, X 'Of Miracles', para. 87

Josselyn, S. A., & Tonegawa, S. (2020). Memory engrams: Recalling the past and imagining the future. *Science*, 367(6473), eaaw4325.

Kleckner, I. R., Zhang, J., Touroutoglou, A., Chanes, L., Xia, C., Simmons, W. K., … & Feldman Barrett, L. (2017). Evidence for a large-scale brain system supporting allostasis and interoception in humans. *Nature Human Behaviour*, 1(5), 0069.

Kumaran, D., Hassabis, D., & McClelland, J. L. (2016). What learning systems do intelligent agents need? Complementary learning systems theory updated. *Trends in Cognitive Sciences*, 20(7), 512–34.

Kurtzman, H. S. (1983). Modern conceptions of memory. *Philosophy and Phenomenological Research*, 44(1), 1–19.

Kwan, D., Craver, C. F., Green, L., Myerson, J., Gao, F., Black, S. E., & Rosenbaum, R. S. (2015). Cueing the personal future to reduce discounting in intertemporal choice: Is episodic prospection necessary?. *Hippocampus*, 25(4), 432–43.

Lanzoni, L., Ravasio, D., Thompson, H., Vatansever, D., Margulies, D., Smallwood, J., & Jefferies, E. (2020). The role of default mode network in semantic cue integration. *NeuroImage*, 219, 117019.

Leal Santos, S., Stackmann, M., Zamora, A. M., Mastrodonato, A., De Landri, A. V., Vaughan, N., & Denny, C. A. (2021). Propranolol decreases fear expression by modulating fear memory traces. *Biological Psychiatry*, 89(12), 1150–61.

Liu, C., Yen, C. C. C., Szczupak, D., Ye, F. Q., Leopold, D. A., & Silva, A. C. (2019). Anatomical and functional investigation of the marmoset default mode network. *Nature Communications*, 10(1), 1975.

Loftus, E. F., & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals*, 25, 720–25.

Long, X., & Zhang, S. J. (2021). A novel somatosensory spatial navigation system outside the hippocampal formation. *Cell Research*, 31(6), 649–63.

Lu, H., Zou, Q., Gu, H., Raichle, M. E., Stein, E. A., & Yang, Y. (2012). Rat brains also have a default mode network. *Proceedings of the National Academy of Sciences*, 109(10), 3979–84.

Lyons, J. C. (2019). Algorithm and parameters: Solving the generality problem for reliabilism. *Philosophical Review*, 128(4), 463–509.

Maguire, E. A., & Mullally, S. L. (2013). The hippocampus: A manifesto for change. *Journal of Experimental Psychology: General*, 142(4), 1180.

Malcolm, N. (1963). *Knowledge and Certainty*. Cornell University Press.

Malcolm, N. (1977). *Memory and Mind.* Cornell University Press.

Marr D. (1971). Simple memory: A theory for archicortex. *Phil. Trans. R. Soc. Lond.*, B 262, 23–81.

Martin, C. B., & Deutscher, M. (1966). Remembering. *Philosophical Review*, 75, 161–96.

McAvan, A. S., Wank, A. A., Rapcsak, S. Z., Grilli, M. D., & Ekstrom, A. D. (2022). Largely intact memory for spatial locations during navigation in an individual with dense amnesia. *Neuropsychologia*, 170, 108225.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102, 419–57.

Michaelian, K., & Robins, S. K. (2018). Beyond the causal theory?: Fifty years after Martin and Deutscher 1. In Michaelian, K., Debus, D., and Perrin, D. *New Directions in the Philosophy of Memory* (pp. 13–32). Routledge.

Michaelian, K. (2016). *Mental Time Travel: Episodic Memory and Our Knowledge of the Personal Past.* MIT Press.

Michaelian, K. (2024). Radicalizing simulationism: Remembering as imagining the (non-personal) past. *Philosophical Psychology*, 37(5), 1170–96.

Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current opinion in neurobiology*, 7(2), 217–27.

Nader, K., & Einarsson, E. Ö. (2010). Memory reconsolidation: An update. *Annals of the New York Academy of Sciences*, 1191(1), 27–41.

O'Keefe, J., & Recce, M. L. (1993). Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus*, 3(3), 317–30.

Parikh, N., Ruzic, L., Stewart, G. W., Spreng, R. N., & De Brigard, F. (2018). What if? Neural activity underlying semantic and episodic counterfactual thinking. *NeuroImage*, 178, 332–45.

Perrin, D. (2018). A case for procedural causality in episodic recollection. In *New Directions in the Philosophy of Memory*. Routledge.

Perrin, D. (2021). Embodied episodic memory: A new case for causalism?. *Intellectica* 74:229–52.

Race, E., Keane, M. M., & Verfaellie, M. (2011). Medial temporal lobe damage causes deficits in episodic memory and episodic future thinking not attributable to deficits in narrative construction. *Journal of Neuroscience*, 31(28), 10262–69.

Race, E., Keane, M. M., & Verfaellie, M. (2013). Living in the moment: Patients with MTL amnesia can richly describe the present despite deficits in past and future thought. *Cortex*, 49(6), 1764–66.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences*, 98(2), 676–82.

Robins, S. K. (2016). Representing the past: Memory traces and the causal theory of memory. *Philosophical Studies*, 173, 2993–3013.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words that were not presented in lists. *JEP:LMC*, 21, 803–14.

Romero, K., & Moscovitch, M. (2012). Episodic memory and event construction in aging and amnesia. *Journal of Memory and Language*, 67(2), 270–84.

Rosenbaum, R. S., Cassidy, B. N., & Herdman, K. A. (2015). Patterns of preserved and impaired spatial memory in developmental amnesia. *Frontiers in Human Neuroscience*, 9, 196.

Rosenbaum, R. S., Priselac, S., Köhler, S., Black, S. E., Gao, F. Q., Nadel, L., & Moscovitch, M. (2000). Remote spatial memory in an amnesic person with extensive bilateral hippocampal lesions. *Nature Neuroscience*, 3(10), 1044–48.

Rosenbaum, R. S., Stuss, D. T., Levine, B., & Tulving, E. (2007). Theory of mind is independent of episodic memory. *Science*, 318, 1257.

Rosenbaum, R. S., Gilboa, A., Levine, B., Winocur, G., & Moscovitch, M. (2009). Amnesia as an impairment of detail generation and binding: Evidence from personal, fictional, and semantic narratives in K.C. *Neuropsychologia*, 47, 2181–87.

Rumelhart, D. E., McClelland, J. L., & PDP Research Group (eds.). (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations*. MIT Press.

Schacter, D. L., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory: Remembering the past and imagining the future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1481), 773–86.

Schacter, D. L., Wagner, A. D., & Buckner, R. L. (2000). Memory Systems of 1999. In E. Tulving & F. I. M. Craik (Eds.), *Handbook of memory*. Oxford University Press.

Schacter, D. L., Benoit, R. G., De Brigard, F., & Szpunar, K. K. (2015). Episodic future thinking and episodic counterfactual thinking: Intersections between memory and decisions. *Neurobiology of Learning and Memory*, 117, 14–21.

Schroeder, M. P., Weiss, C., Procissi, D., Disterhoft, J. F., & Wang, L. (2016). Intrinsic connectivity of neural networks in the awake rabbit. *Neuroimage*, 129, 260–67.

Sforazzini, F., Schwarz, A. J., Galbusera, A., Bifone, A., & Gozzi, A. (2014). Distributed BOLD and CBV-weighted resting-state networks in the mouse brain. *Neuroimage*, 87, 403–15.

Skaggs, W. E., McNaughton, B. L., Wilson, M. A., & Barnes, C. A. (1996). Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus*, 6(2), 149–72.

Spreng, R. N., Mar, R. A., & Kim, A. S. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. *Journal of cognitive neuroscience*, 21(3), 489-510.

Spreng, R. N., & Andrews-Hanna, J. R. (2015). The default network and social cognition. *Brain Mapping: An Encyclopedic Reference*, 1316, 165–69.

Szpunar, K. K., Watson, J. M., & McDermott, K. B. (2007). Neural substrates of envisioning the future. *Proceedings of the National Academy of Sciences*, 104(2), 642–47.

Talland, G. A. (1965). *Deranged Memory: A Psychonomic Study of the Amnesic Syndrome.* Academic Press.

Teyler T. J., & DiScenna, P. (1986). The hippocampal memory indexing theory. *Behavioral Neuroscience*, 100, 147–52.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, 26, 1–12.

Van Hoeck, N., Ma, N., Ampe, L., Baetens, K., Vandekerckhove, M., & Van Overwalle, F. (2013). Counterfactual thinking: An fMRI study on changing the past for a better future. *Social Cognitive and Affective Neuroscience*, 8(5), 556–64.

Vincent, J. L., Patel, G. H., Fox, M. D., Snyder, A. Z., Baker, J. T., Van Essen, D. C., & Raichle, M. E. (2007). Intrinsic functional architecture in the anaesthetized monkey brain. *Nature*, 447(7140), 83–86.

Vosgerau, G. (2010). Memory and content. *Consciousness and Cognition*, 19(3), 838–46.

Werning, M. (2020). Predicting the past from minimal traces: episodic memory and its distinction from imagination and preservation. *Review of Philosophy and Psychology*, 11(2), 301–33.

Wikenheiser, A. M., Gardner, M. P., Mueller, L. E., & Schoenbaum, G. (2021). Spatial representations in rat orbitofrontal cortex. *Journal of Neuroscience*, 41(32), 6933–45.

Zhang, R., & Volkow, N. D. (2019). Brain default-mode network dysfunction in addiction. *Neuroimage*, 200, 313–31.

PART III

10

Can We Perceive the Past?

E. J. Green

Introduction

What distinguishes perception from memory? A natural answer is that perception tells us about the present, while memory tells us about the past. When you see a red tomato on your kitchen table, your perceptual state represents that there is *now* a red tomato on the table. Conversely, your perceptually based memory of the tomato represents that there *was* a red tomato on the table.

By *perceptual presentism*, I will mean, very roughly, the view that perception is restricted to representing present conditions. Perceptual presentism has a long intellectual history. In the *Confessions*, Augustine writes: 'When time is passing, it may be perceived and measured; but when it is past, it cannot, because it is not' (Gale 1968: 42). And Thomas Reid writes:

> The object of memory, or thing remembered, must be something that is past; as the object of perception and of consciousness must be something which is present: What now is, cannot be an object of memory; neither can that which is past and gone be an object of perception or of consciousness. (Reid 1785/2011: essay III, ch. I)[1]

Perceptual presentism also has important implications. For example, it is often held that perception is epistemically distinctive insofar as having a perceptual experience that represents that p provides immediate, non-inferential justification for believing that p (Pryor 2000; Silins 2021). But if perception is restricted to representing present conditions, then this route to non-inferential justification only covers beliefs about the present. Past- or future-oriented beliefs must be justified via inference, or non-inferentially via some other route.

Moreover, several have suggested that a fundamental function of perception is to deliver 'news' to the rest of the mind, while cognition functions to draw inferences from that news (Beck 2018; Block 2023). This view may seem to support perceptual presentism. Intuitively, events occurring *now* are news, while past events aren't. As Block (2023) writes: 'Perception functions to provide us with information about

[1] Contemporary theorists expressing sympathy with perceptual presentism include Peacocke (1999: 280), Carey (2009: 9), Kriegel (2015), Gross (2017), Beck (2018), Byrne (2018: ch. 8), Hoerl (2018), Connor and Smith (2019), and Block (2023: 20, 119).

E. J. Green, *Can We Perceive the Past?*. In: *Space, Time, and Memory*. Edited by: Lynn Nadel and Sara Aronowitz, Oxford University Press. DOI: 10.1093/oso/9780192882547.003.0010

what is happening in the nearby environment now, whereas cognition functions in reasoning about the news provided by perception so as to decide what to do and to plan for the future' (20). (I'll return to this idea in section 5a., however, where I'll argue that the news-propagating function of perception does not provide strong support for perceptual presentism.)

Furthermore, some have held that what differentiates higher cognitive faculties distinctive of human thought from 'lower' faculties, like perception, shared with non-human animals is the ability to 'mentally time travel' to contemplate scenarios in the past or future (Suddendorf & Corballis 2007). This hypothesis comports with the view that perceptual representations are locked to the present, while higher cognition is more temporally flexible. However, if representation of the past or future occurs even in perception, this would challenge proponents of the hypothesis to explain how (or whether) the temporal flexibility of perception differs from the temporal flexibility of thought.

While perceptual presentism is widely assumed, a substantial body of evidence indicates that perception is sensitive to the past in intricate ways. An object's appearance is determined not only by present sensory stimulation, but by stimulation in the recent and distant past. Theorists moved by such evidence occasionally venture claims that seem to conflict with perceptual presentism. Thus, Munton (2022) claims that 'representations from memory are intimately woven into perceptual representations' (2), and that 'far from living in the present, an entwinement with the past is inherent in the most basic of perceptual capacities, our perception of objects' (18). Similarly, Pascucci et al. (2019) write that 'the content of visual perception is strongly permeated by information lingering from the past' (20). But what is the nature of this 'entwinement' or 'permeation' between perception and the past? And does it threaten perceptual presentism?

This chapter critically examines perceptual presentism in light of perception science. I don't aim to conclusively establish or refute the view. Rather, I aim to pinpoint the sort of discovery that would imperil perceptual presentism, and to filter out forms of evidence that do not. I distinguish three forms of perceptual sensitivity to the past: (i) shaping perception by past stimulus exposure, (ii) recruitment of mnemonic representations in perceptual processing, and (iii) perceptual representation of present objects as possessing past properties. I'll adduce evidence for each. I'll argue that forms (i) and (ii) are likely consistent with perceptual presentism, while form (iii) poses a genuine threat to the view. While the case for this form of sensitivity remains inconclusive, I suggest that the most compelling challenges to perceptual presentism likely derive from representations that seamlessly integrate mnemonic and present-tensed elements in the performance of canonical perceptual functions, such as apprehending object continuity over time.

Two caveats. First, due to space limitations, I will only discuss perceptual sensitivity to the *past*. Certain kinds of perceptual sensitivity to the *future*—e.g., prediction or motion extrapolation (Clark 2015; Hogendoorn 2020)—might be construed as involving future-tensed perceptual representation. Such cases are obviously also relevant to evaluating perceptual presentism. However, they raise further issues that

must be considered independently (White 2018). Second, my discussion will be regrettably limited to vision. Non-visual capacities clearly also matter for evaluating perceptual presentism, and some such capacities (say, hearing a melody as extending over a long interval) might be thought to put serious pressure on presentism (e.g., Viera 2022: 639). Note, for present purposes, that if perceptual presentism fails for non-visual modalities, supporters of the view might retreat to a vision-restricted version of the thesis.

Formulating Perceptual Presentism

Taken literally, Reid's claim that the 'object of perception and of consciousness must be something which is present' faces obvious counterexamples. Arguably, when we see a distant star that has ceased to exist, *what* we perceive—the *object* of perception—is something past, not present. Ergo, not all objects of perception are present.

However, one can accommodate these classic 'time-lag' cases while preserving the spirit of perceptual presentism. Rather than imposing a presentist restriction on the *things* (objects or events) we perceive, perceptual presentism can be understood as a view about how perception *represents* the things we perceive: when an object, event, or state of affairs is represented in perception, it is always represented *as present*—i.e., as presently located or instantiated in one's surroundings—even if it turns out not to be present.[2] Thus, perception is, or at least purports to be, wholly about what's going on now.

There are various forms that that this presentist restriction on perceptual representation might take (Skow 2011; Kriegel 2015; Connor & Smith 2019). However, here I'll focus on a version of the view that builds the presentist restriction into the *content* of perception. Specifically, for any individual O and property F, if a perceptual state S attributes F to O, then S represents that O is F *now*, where 'now' denotes either the time of S's occurrence or a brief interval containing that time (Peacocke 1999: 280).[3] When your perceptual state attributes brownness to a table, it represents that the table is brown *now*.

Some philosophers might be uncomfortable with the idea that perception has genuinely tensed content. For example, those who claim that perception is wholly iconic or image-like (e.g., Block 2023) might argue that perceptual representations

[2] Barkasi (2021: 7) likewise argues that memory introduces past properties into perception, but denies that these properties are perceptually represented *as* past.

[3] Alternatively, one might pack presentness into the *attitude* that perceptual states bear to their contents: Roughly, whenever a perceptual state S attributes a property F to an individual O, S *represents-as-present* that O is F, and is accurate only if O is F at the time of S (see Kriegel 2015). I believe that most of my points could be adjusted to apply to attitude-based versions of perceptual presentism, but I won't attempt the needed adjustments here. See also Connor and Smith (2019) for further ways of developing perceptual presentism, and Hoerl (2018) for an attempt to characterize the relation between perception and the present without invoking any 'temporal mode of presentation' within perception (129).

lack the sort of syntactic complexity required for explicit tense markers like tensed copula. Two points on this. First, it is an empirical question whether perception is fully iconic, and the proposal has faced significant pushback. Just as perceptual representations arguably possess discrete, non-iconic singular constituents (Green 2023), they may also possess discrete, non-iconic tense markers. Second, proponents of iconicity might grant that perceptual representation is tensed without conceding that perceptual representations contain discrete tense markers. Instead, they might possess tensed content in virtue of their functional role (compare Block 2023: ch. 5). If consuming processes are reliably disposed to *treat* a perceptual representation as if its content is present-tensed, this might simply make it the case that it has present-tensed content. (A parallel story could perhaps apply to past-tensed perceptual representations, if there are any—see section 5b.)

I am construing perceptual presentism as a thesis about perceptual representation in general. Some might wish to restrict the thesis to *conscious perceptual experience*. Indeed, some might think that we can establish presentism about perceptual experience through introspection alone: When we reflect on perceptual experience, it always *seems to us* that the experience simply makes us aware of how the environment is presently.

This issue deserves more attention than I can give it here. For now, I simply register my doubts about any quick introspective argument for perceptual presentism. These doubts primarily stem from examples of the sort to be considered in section 5a. When you see a bitten cookie (Figure 10.1), there *does* seem to be an immediate, conscious impression of a 'past' aspect of the object—that it *was* circular, but no longer is. It's possible that this impression is wholly post-perceptual, but I submit that it is not *obvious*, introspectively, that this is so.

Moreover, while it could conceivably turn out that perceptual presentism is true of conscious perception but false of perception more generally, we would then like to know what prevents the non-present-tensed elements of perceptual representation from *becoming* conscious. Thus, those who are mainly concerned with conscious perception should still find value in examining presentism about perception more generally.

Perceptual presentism claims that perceptual states purport to be wholly about the present. But what, more precisely, do we mean by the 'present'? A familiar debate concerns whether the perceived present is a durationless instant or an extended interval—a *specious* present (James 1890; Phillips 2011, 2014; Dainton 2022). Specious-present theorists hold that perceptual awareness of dynamic phenomena like motion, persistence, and change shows that the fundamental 'units' of perceptual experience represent extended intervals. The intervals must be long enough for perceptible instances of motion or change to unfold—commonly estimated at 100–500 ms in length (Phillips 2011, 398; White 2020; Herzog et al. 2020). Opponents of specious-present theory have argued that motion, change, etc. are not genuinely perceived, only cognized (Chuard 2011).

While I cannot defend it here, I agree with specious-present theorists that the perceived present is extended.[4] I suggest that when you perceptually represent <there is *now* a red tomato thereabouts>, the indexical 'now' denotes an interval roughly simultaneous with the perceptual state, but potentially including moments shortly before or after the state's onset. As White (2020) observes, the perceived present plausibly comprises an interval longer than the minimum threshold for discriminating the temporal order of events (roughly 20 ms (White 2020, 587)). One event might be perceived as preceding another, even though both are marked 'present' (2020, 587–88).

Summing up: Perceptual presentism claims that when a perceptual state attributes a property F to an object O, its accuracy depends entirely on whether O has F presently. However, the 'present' can be taken to encompass an interval rather than a durationless instant. Thus, perceptual presentism is compatible with basic perceptual awareness of motion and change.

Shaping Perception by Past Stimulus Exposure

The first kind of perceptual sensitivity to the past involves shaping perceptual processing by past stimulus exposure. It is well established that perceptual processing at a time is causally influenced by stimuli encountered seconds, minutes, or even months prior. I mention three examples.

First, in *perceptual learning*, training in a perceptual task modifies perceptual processing to facilitate performance of the task (Goldstone 1998). Studies have documented learning-induced changes in responses at early stages of sensory processing (Schoups et al. 2001). Thus, Yan et al. (2014) found that after extended training in a contour detection task, neurons in macaque primary visual cortex began to fire more vigorously when an edge in their receptive field fell along an extended contour. The magnitude of this physiological change correlated with improvements in behavioural performance. Further evidence from functional magnetic resonance imaging (fMRI) (Furmanski et al. 2004; Jehee et al. 2012) and electroencephalogram (EEG) (Pourtois et al. 2008; Bao et al. 2010) studies suggests that perceptual learning modifies early visual cortical responses in humans too.[5]

Second, in *adaptation*, exposure to a stimulus biases perception of a subsequent stimulus 'away' from the first. If you view an upward-moving pattern for thirty

[4] Among specious-present theorists, there is a further distinction between *retentional* models, on which experiences are instantaneous (or very brief) events that *represent* a longer interval (Herzog et al. 2020), and *extensional* models, on which experiences, or at least the 'metaphysically fundamental units' of experience (Phillips 2011: 398), are temporally extended events lasting as long as a specious present (Phillips 2011; Dainton 2022). I will remain neutral on this issue.

[5] There is controversy about whether these changes reflect feedback from higher areas, and whether they involve signal enhancement or internal noise reduction (Dosher & Lu 2017). Moreover, some studies fail to find an impact of perceptual learning on early sensory areas (Law and Gold 2008). Still, the dominant view seems to be that perceptual learning alters perceptual processing, 'tuning' our perceptual systems for certain sorts of detection or discrimination (Connolly 2019: ch. 2).

seconds, a stationary pattern shown immediately afterward appears to move downward (Anstis et al. 1998). Adaptation after-effects occur not only for low-level properties (motion, colour, etc.), but for high-level properties like facial expressions (Butler et al. 2008), causation (Rolfs et al. 2013; Kominsky & Scholl 2020), and numerosity (Burr & Ross 2008). While most after-effects are short-lived, some can persist for months after stimulus exposure (Jones & Holding 1975; Thompson & Burr 2009: R13).

The third example is *serial dependence*. While adaptation effects are typically *repulsive*—they bias perception away from a recently perceived feature—serial dependence (SD) involves *attractive* effects of past stimuli on perception. Fischer & Whitney (2014) found that when subjects were repeatedly shown Gabor patches and asked to indicate their orientation, responses were biased towards the orientations of patches seen within the past two to three trials (roughly ten seconds). SD is partially spatially tuned—attraction is stronger between stimuli presented in the same location (Collins 2019). While adaptation and SD may seem like conflicting phenomena, adaptation seems to require longer stimulus exposure than SD (Fischer & Whitney 2014), and SD requires attention while adaptation does not (Fritsche & de Lange 2019).

I should flag that it is controversial whether SD arises in perception or post-perceptual decision processes (Fritsche et al. 2017; Pascucci et al. 2019; Ceylan et al. 2021). Some suggest that SD *originates* at post-perceptual levels (i.e., in high-level decisions about stimuli) but *affects* perceptual representation of later stimuli (Pascucci et al. 2019: 22; Ceylan et al. 2021; Phillips & Firestone, forthcoming). If SD is purely post-perceptual in both its origin and effects, then it obviously poses no challenge to perceptual presentism. For present purposes, I'll assume that at least some instances of SD are perceptual, since I don't believe it significantly threatens perceptual presentism either way.

Perceptual learning, adaptation, and SD demonstrate rich influences of past stimuli on perception. However, these phenomena do not yet put serious pressure on perceptual presentism. The problem is that while they involve *effects* of past stimuli on perceptual representation, they do not require those past stimuli to be *explicitly represented* when their effects materialize.

When perceptual processing is sensitive to a piece of information, there are two ways this sensitivity could be achieved. First, the information might be explicitly represented in the perceptual system and retrieved while computing distal features from proximal stimuli. Second, the information might be 'implicitly embodied' in the system's dispositions to transition between information-carrying or representational states (Pylyshyn 2003; Shea 2015). To take a familiar example, the visual system's assumption that light comes from above might be explicitly represented, or it might be implicit in a disposition to transition from encodings of luminance patterns on the retina to representations of surface convexity.

The distinction between explicitly represented and implicitly embodied information is not entirely clear-cut. However, one prototypical difference is that explicit

representations are made available to a variety of consuming processes, while a disposition to transition between information-carrying or representational states is not (Shea 2015; Clark 1992). If the assumption that light comes from above is implicitly embodied in a disposition to transition from registrations of luminance patterns to representations of surface convexity, then the assumption is only 'available' to the process of computing convexity. It cannot be freely accessed by other computations.

Perceptual learning, adaptation, and SD showcase perception's exquisite sensitivity to past stimuli, but do not obviously require explicit representations of those past stimuli. Recall Yan et al.'s (2014) finding that training in a contour detection task led neurons with receptive fields along an extended contour to fire more vigorously. Arguably, past training trials affected the visual system's disposition to transition from states encoding local edge elements to states encoding (crudely) that an edge element at some retinotopic location lies along an extended contour. But we needn't posit an explicit representation of these past trials, since (inter alia) there is no evidence that information about them is available to processes outside contour detection.

Likewise, classic examples of adaptation don't seem to require explicit representation of the past stimuli inducing the after-effect. Adaptation generally occurs when the perceptual system codes for a dimension, such as motion or colour, via ratios of activity in multiple channels. Extended exposure to a stimulus reduces the sensitivity of certain channels more than others, shifting the activity ratio elicited by a subsequent stimulus away from what it would normally be (Anstis et al. 1998; Webster 2011; Block 2023: ch. 2). This is a shift in state–transition dispositions. For example, the visual system has dispositions to transition from motion energy signals at the retina to representations of motion. These dispositions are grounded in the sensitivities of various motion-coding channels. Tweaking these sensitivities through adaptation lowers the threshold for representing either upward or downward motion. However, at the time of the after-effect, there need be no explicit representation of the adaptation stimulus.

The case of SD is more complicated because it is less understood. However, there is evidence that at least some cases of SD do not require explicit mnemonic representation of the earlier stimulus exerting an attractive effect. (Other cases of SD may be mediated by explicit short-term memory representations (Czoschke et al. 2019). Such cases would fall under the second form of perceptual sensitivity to the past, discussed in section 4.)

In one experiment, Fischer and Whitney (2014) asked participants to report both the orientation of a currently perceived Gabor patch, and whether the stimulus shown two trials earlier was oriented clockwise or counterclockwise from vertical. They found that subjects were virtually at chance in reporting the orientation of the earlier stimulus. However, that stimulus still had an attractive influence on reports about the later stimulus. Moreover, when orientation was remembered inaccurately, attraction tended towards the *actually displayed* orientation, not the *inaccurately remembered* orientation. Arguably, then, SD can occur without explicit mnemonic

representation. In such cases, SD may depend on modifications to state–transition dispositions.

Suppose, then, that perceptual learning, adaptation, and certain cases of serial dependence are wholly explained by changes to state–transition dispositions within our perceptual systems. Do these phenomena threaten perceptual presentism? I think the answer is no.

First, while some are happy to speak of information embodied in a system's state–transition dispositions as 'implicitly represented' (Shea 2015), others argue that there is no reason to hold that such information is represented at all (Ramsey 2007: ch. 5). Suffice it to say that *if* implicitly embodied information is not represented in *any* sense, then implicit embodiment of information about the past within our perceptual systems poses no threat to perceptual presentism.

Suppose, however, that it is reasonable to speak of implicit representation in these cases. Still, perceptual presentism is naturally understood as a view about what perception *explicitly* represents. It is a view (inter alia) about the content that perception *makes available* to other mental processes, such as belief fixation and planning. Perceptual presentists hold that all contents 'given' to us in perception concern the present. But information embodied in state–transition dispositions is not given to us in this sense. It is available only to certain subpersonal processes within our perceptual systems. Phenomenological reflection bolsters this point. When we undergo after-effects, such as afterimages, the afterimage *seems like* it is out there in the present environment—it doesn't strike us as past. Thus, even if information about the past is implicitly represented by state–transition dispositions within perceptual systems, perceptual presentists can reasonably reply that their view is restricted to explicit perceptual representation.

Mnemonic Representations in Perceptual Processing

I've considered cases in which perception is influenced by past factors but there is no good reason to accept that those past factors are explicitly represented. I now turn to cases where there *is* evidence that past factors are explicitly represented and recruited in perceptual processing. Unlike adaptation and perceptual learning, these cases arguably support the presence of explicit past-tensed representations within our perceptual systems.

Visual working memory (VWM) is a limited-capacity information store that can retain representations of past stimuli for several seconds after they disappear (Brady et al. 2011; Suchow et al. 2014). Importantly, there is evidence that representations retained in VWM can coherently affect perceptual processing.

First, VWM recruits some of the same brain areas as perception. Physiological studies indicate that when an item is retained in VWM, information about the item can be decoded from population-level activity in early visual cortex (Harrison & Tong 2009; Serences et al. 2009). However, while such evidence is suggestive, it

doesn't establish that VWM representations are accessed during perceptual processing. VWM might use some of the same brain areas as perception without playing a computational role in perception. Nonetheless, behavioural evidence supports an active role for VWM representations in perceptual processing.

Kang et al. (2011) found that VWM coherently affects motion perception. The *motion repulsion effect* occurs when subjects are shown superimposed sets of dots moving in different directions, and the perceived angular separation between their directions of motion is exaggerated relative to their true separation. Kang et al. found that simply retaining a motion stimulus in VWM can produce a motion repulsion effect. Participants saw two 500-ms dot motion displays separated by a two-second interval. After the second display's offset, they indicated whether its direction of motion was clockwise or counterclockwise relative to a reference bar. Critically, when the first display was retained in VWM, the second display elicited a repulsion effect: subjects' reports indicated that perceived direction of the second stimulus was repelled from the direction of the first. Conversely, when subjects made immediate reports about both stimuli without a recall task, no repulsion effect emerged. Assuming that these reports reflected participants' perceptual experiences, we can conclude that retaining the first stimulus in memory affected perception of the second stimulus. There is also evidence that VWM influences perception of ambiguous motion stimuli (Scocchia et al. 2013; Hein et al. 2021), orientation, and colour (Teng & Kravitz 2019).

There are many other proposed effects of memory on perception. Munton (2022) appeals to the role of long-term memory in perception, including the recruitment of Bayesian priors (Weiss et al. 2002; Kersten et al. 2004). She also cites the phenomenon of *boundary extension*, where the representation of a scene depicted in a picture extends beyond the boundaries of the picture (Intraub & Dickinson 2008; Hubbard et al. 2010). However, these cases face certain interpretive difficulties. It is controversial whether Bayesian priors are explicitly encoded in perception, or are merely implicit in the visual system's state–transition dispositions (Orlandi 2016; Sanborn & Chater 2016; Icard 2016; Block 2018; Rescorla 2020). And as Munton observes, it is unclear whether boundary extension is genuinely an effect on perceptual representation, or arises while encoding perceptually represented information into memory (or perhaps involves perceptually based imagery—Nanay 2021).[6]

I focus on VWM because I believe it offers the strongest argument for the use of *explicit past-tensed representations* in perceptual processing. When information is retained in VWM, it is made accessible for retrieval and report in recall tasks. Such wide accessibility is a hallmark of explicit representation. Furthermore, it's at least plausible that when a representation of an object is retained in VWM, the object is represented or 'tagged' *as* past. Such tagging would allow VWM and perceptual representations to be integrated in reasoning without creating confusion about

[6] Or perhaps boundary extension emerges only when *reporting* remembered scene boundaries, not in either perception or memory encoding.

which things are currently there and which disappeared several seconds ago (White 2020: 594). By contrast, higher-capacity sensory memory stores (e.g., iconic or fragile short-term memory—Sperling 1960; Landman et al. 2003) are more short-lived than VWM, making temporal tagging less critical.

Suppose, then, that perceptual processing is sometimes guided by mnemonic representations with past-tensed content. Does this immediately threaten perceptual presentism? I believe the answer is no.

Recall Munton's (2022) claim that 'an entwinement with the past is inherent in the most basic of perceptual capacities' (18). It is important to distinguish *causal* from *constitutive* versions of this claim. On one view, perception is entwined with the past just insofar as mnemonic representations causally influence the production of perceptual representations. On another view, perception is entwined with the past insofar as some perceptual representations are either partially or fully constituted by mnemonic representations (i.e., have such representations as constituents).[7] The foregoing evidence only directly supports the causal claim: representations retained in VWM influence the production of perceptual representations (e.g., of motion).[8] However, the constitutive claim is the relevant one for evaluating perceptual presentism.

Does the fact that mnemonic representations are computed over in perceptual processing suffice to show that they are perceptual representations? If so, then there would be less distance between the causal and constitutive claims than it first appears, since the right sort of causal involvement in perceptual processing would suffice to make a mnemonic representation perceptual, establishing the constitutive claim.

The answer depends on what we mean by a 'perceptual representation'. Perceptual presentism claims that perceptual representations have entirely present-tensed content. But what *makes* a mental representation 'perceptual', as opposed to doxastic or something else?

On one view, perceptual representations are simply representations 'in' one's perceptual system. They are representations within the database to which the perceptual system has access during computation. Call this the *system view* of perceptual representation—so called because it distinguishes perceptual from non-perceptual representations by appeal to the systems that use them. The view needn't claim that being a perceptual representation *excludes* being a representation of another type. If a state is accessed by systems of both perception and action, then it might count as both a perceptual representation and a motor representation.

If the system view were right, then the foregoing evidence arguably would refute perceptual presentism. For it demonstrates that VWM representations are within

[7] However, even if perceptual representations are partially constituted by mnemonic representations, then it is still a further question whether the latter representations are genuinely *past-tensed* (see section 5b).

[8] Munton is primarily concerned to argue that mnemonic information informs the perceptual representation of objects while they are occluded, and thus allows us to 'see' invisible objects. Note that one could grant that memory affects perception in this way without granting that mnemonic representations partially constitute perceptual representations.

the database of representations that our perceptual systems can access when computing representations of distal conditions. If VWM representations are genuinely past-tensed, then we must conclude that there are perceptual states with past-tensed content, and so perceptual presentism is false.

However, perceptual presentists should reject the system view, since it provides a too permissive criterion of perceptual representation. It is highly implausible that *any* mental state within the stock of representations accessed during perceptual processing is thereby a perceptual representation. Consider cognitive penetration. It has been argued that desires can influence distance perception, whereby desired objects appear closer than non-desired objects (Balcetis & Dunning 2010). While there are reasons to doubt this claim (Durgin et al. 2011; Firestone & Scholl 2016), suppose for the moment that it is true, *and* that the effect involves direct, unmediated access to desires during perceptual processing (cf. Macpherson 2012). Then desires would be within the database to which perceptual processes have access. Still, it would seem wrong to conclude that *desires* are therefore *perceptual representations*.

So, there are reasons to regard the system view as too permissive.[9] But this does not yet constitute an argument *against* treating offline VWM representations as perceptual representations when they are accessed during perceptual processing. I now consider a more restrictive criterion that may serve this purpose.

Various authors have suggested that perceptual states are *stimulus-dependent* in a way that cognitive states are not (Beck 2018; Phillips 2019; compare Rock 1982; Carey 2009: 9). Roughly, perceptual states function to be produced by proximal stimulation of the sense organs, and to be updated in response to changes in proximal stimulation. If a state is completely untethered from proximal stimulation, then either it is non-perceptual or it must involve some malfunction of the sensory system (e.g., in certain cases of hallucination).

Beck (2018) endorses the stronger claim that, necessarily, *all constituents* of perceptual states function to be stimulus-dependent. Roughly, any perceptual attributive—any element of a perceptual state that indicates a property—functions to be causally sustained by some relevant aspect of proximal stimulation—i.e., a 'cue' to the property it represents. However, this claim has faced pushback (Quilty-Dunn 2020a). Moreover, cases of perceptual tracking through full occlusion raise prima facie difficulties for strong stimulus-dependence. Perceptual object representations are plausibly sustained through periods of occlusion where no proximal cues are received from the object (Scholl & Pylyshyn 1999; Flombaum et al. 2008; Munton 2022). Here, a perceptual representation of the object, and arguably attributives for at least some of its features, e.g. shape or size, are sustained not by proximal cues supplied by the object or its features, but (if anything) by proximal stimulation supplied by the occluder.

[9] Gross (2017) likewise claims that there can be representational states within perceptual systems—viz., attentional commands—that are not perceptual representations. However, his basis for denying that attentional commands are perceptual representations seems to assume perceptual presentism: 'Consider the … attentional command. Though it is perhaps (if we deny it cognitive status) a representational state *in* perception, it is not itself a perceptual state, at least in the sense of a state whose function is to represent the here and now' (6).

I espouse only *modest* stimulus-dependence criterion as a signature mark of perception. Canonically, perceptual representations (i) function to be produced by proximal stimulation, and (ii) function to be updated in response to changes in proximal stimulation.[10] When your perceptual system is functioning normally, a perceptual representation of a red apple is generated in response to proximal stimulation received from the apple, and the representation is disposed to be updated in response to changes in proximal stimulation. Notice that object and feature representations retained through occlusion meet this modest stimulus-dependence criterion. The representations are generated in response to proximal stimulation, and are disposed to be updated in response to relevant proximal changes (e.g., if the object emerges from occlusion and has changed colour).

Importantly, many VWM representations do not meet even this modest stimulus-dependence criterion. VWM representations can be retained 'offline', where they function to remain unchanged regardless of changes in proximal stimulation. VWM representations *may* be altered by incoming proximal stimulation due to interference from perceptual processes in overlapping brain areas (Teng & Kravitz 2019; Adam et al. 2021), but plausibly they do not *function* to be altered this way. Their being so altered does not produce any reliable adaptive benefit for the organism. This offers a principled reason to withhold perceptual status from them.[11,12]

One might argue that appealing to stimulus-dependence is question-begging in this context. For, if perceptual presentism trivially *follows* from stimulus-dependence, then anyone sceptical of perceptual presentism would simply reject the stimulus-dependence criterion. However, perceptual presentism does not trivially follow from stimulus-dependence, so this concern is misplaced. There is no inconsistency in the idea that representations of past properties, states, or events might function to be produced and updated in response to proximal stimulation. In fact, I'll argue in the next section that some of the most interesting challenges to perceptual presentism are cases of just this sort.

Summing up: there is evidence that mnemonic representations are recruited during perceptual processing. However, assuming modest stimulus-dependence as a signature marker of perception, we can exclude 'offline' mnemonic representations from the realm of perceptual states, even when they causally influence perceptual processing.

[10] I do not offer this as a strictly necessary or sufficient condition for perception, or as a general theory of the perception-cognition border (for my views on the latter, see Green (2020)).

[11] My point is not that mnemonic representations can *never* be perceptual representations. Rather, it is that *some* mnemonic representations—those that lack even a modest form of stimulus-dependence—are plausibly not perceptual. I return to this issue in section 5b.

[12] While modest stimulus-dependence provides *one* way of denying perceptual status to offline VWM representations, it may not be the only way. Theorists have proffered various empirically necessary conditions on perception, such as iconic format (Block 2023), modularity (Mandelbaum 2018; Quilty-Dunn 2020b), or dimension-restriction (Green 2020, 2023). Perhaps offline VWM representations lack one or more of these features. I emphasize stimulus-dependence because I think it offers the most straightforward rationale against treating offline VWM representations as perceptual, and most theorists would agree that it is at least a reliable signature of perception.

Past Properties of Present Objects

I turn to the cases I take to pose the most serious challenge to perceptual presentism—the perception of present *objects* as having past *properties*. Past-tensed representations may feature as constituents of more complex perceptual representations that also possess present-tensed content: e.g., <O *is* F, but *was* G>. Various cases might be thought to work this way. I'll suggest a general strategy on behalf of perceptual presentists for handling them, which I call the *complexity gambit*. I'll consider an example where the complexity gambit is viable, and then an example where it is not.

Causal History from Shape

Consider the bitten cookie in Figure 10.1. You probably have an irresistible impression that the cookie *used to be* roughly circular before it had a portion removed. That is, you have an impression of its 'causal history'. Some have suggested that an object's causal history might be extracted during perceptual processing on the basis of cues in the proximal stimulus (Leyton 1989, 1992; Spröte et al. 2016; Chen & Scholl 2016). If so, a natural interpretation is that you enjoy a perceptual representation with partly past-tensed content. Call the cookie's precise bitten shape 'B'. Then the representation is something like: <That object is *now* B, but *was* circular>.

Figure 10.1 A bitten cookie.

Source: Spröte and Fleming (2016). Reprinted with permission of Elsevier.

If this is the correct account, then it poses a more difficult problem for perceptual presentism than the mere recruitment of mnemonic representations in perceptual processing. For, the representation of the cookie as formerly circular might be just as stimulus-dependent as the representation of its current shape. The representation of circularity is not like a VWM representation retained offline while the cookie is absent. Rather, like the representation of the cookie's current shape, the representation is formed on the basis of shape-relevant cues in the proximal stimulation

received from the cookie. It might also be updated in response to proximal changes. For example, if the cookie were further distorted (e.g., crumbled) to remove any hint of its past circularity, the representation of circularity might be discarded.

Earlier, I mentioned the idea that a fundamental purpose of perception is to deliver 'news' to the rest of the mind on the basis of sensory stimulation. While that idea might appear to support perceptual presentism (as Block (2023: 20) seems to suggest), we can now see why this is not quite right. For perception to provide 'news', it only needs to deliver information about the environment that is new *for the perceiver*. But past-tensed content can surely qualify as news in this sense. If I enter my kitchen and see a bitten cookie, then it is news to me that the kitchen contains a formerly circular cookie. For certain purposes, such as monitoring my daughter's sugar consumption, this news might be rather important.

Roland Fleming's group has performed numerous studies of the recovery of causal history from shape. These studies probe our ability to judge which shape something had before undergoing some distortion, and to classify shapes by the particular kind of distortion they have undergone (Schmidt & Fleming 2016; Spröte & Fleming 2016; Schmidt & Fleming 2018; Schmidt et al. 2019). In many cases, the relative contributions of perception and post-perceptual cognition in generating subjects' responses are unclear (as the authors acknowledge—Schmidt et al. (2019, 168)). However, I'll highlight the study that I think bears most directly on the issue of whether recovery of causal history is genuinely perceptual.[13]

Spröte, Schmidt, and Fleming (2016) created shapes that appeared either 'complete' or 'bitten' and asked participants to mark the perceived axis of symmetry of the shape by placing dots along that axis. For complete shapes, subjects' responses clustered around the real symmetry axis, while for bitten shapes, responses instead clustered around the axis of symmetry of its *unbitten counterpart* (Figure 10.2). Parallel results emerged for judgements about the front, back, and centre of the shapes. Spröte et al. take these results to suggest that the visual system constructs representations of both the present and past shapes of a bitten object. But presumably we do not perceptually represent these objects as *currently* possessing their past shapes—we are not confused about whether the cookie in Figure 10.1 is presently circular or bitten. The obvious alternative is that pre-bite shapes are attributed in the past tense, while post-bite shapes are attributed in the present tense.

One might suggest that representation of the completed shape did not occur in perception but only in cognition, and these cognitive judgements guided performance on the dot placement task. However, if completed shapes affect the representation of symmetry axes, then this is some evidence that they are represented perceptually. There is evidence that the perceptual system recovers axes of global symmetry and elongation and uses them in constructing visual shape representations (Humphreys & Quinlan 1988; Quinlan & Humphreys 1993; Chaisilprungraung

[13] Chen and Scholl (2016) performed another study sometimes taken to demonstrate that causal history is genuinely perceived. However, this study probed the perception of dynamic transformations (viz., intrusions) within a brief apparent motion sequence—i.e., *very recent* causal history. In this case, it is plausible that causal transformations were represented wholly within the bounds of a single specious present. Given the qualification in section 2, this would be consistent with perceptual presentism.

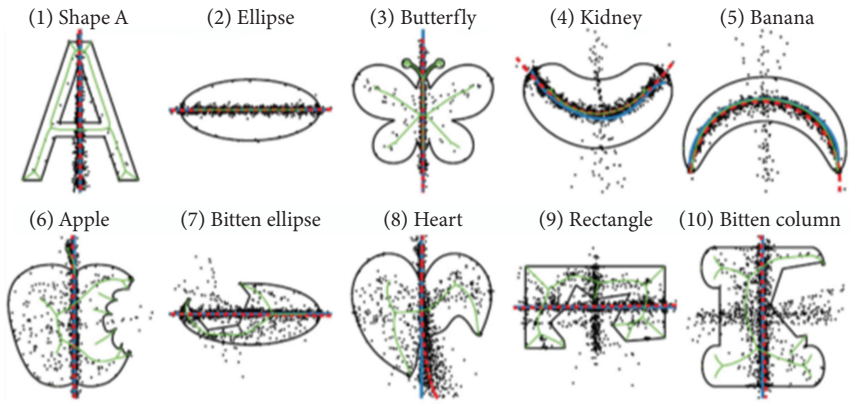


Figure 10.2 Stimuli used by Spröte et al. (2016).

Source: Spröte et al. (2016). Reprinted under the terms of the Creative Commons CC BY license.

et al. 2019). If the processes that recover completed shapes causally influence the processes responsible for recovering symmetry axes, this lends defeasible support to the view that the former processes are perceptual too (see also Chen & Scholl 2016).

Rather than dwell on the perception-cognition issue, I will assume for the sake of argument that completed, pre-bite shapes *are* perceptually represented. My question is whether, granting this, we must conclude that completed shapes of bitten objects are perceptually attributed in the *past tense*. I think the answer is no: perceptual presentists can accept that completed shapes are perceptually represented while denying that they are attributed in the past tense.

The most flatfooted strategy would be to hold that perception of bitten shapes involves a familiar kind of perceptual completion. Perceptual completion comes in two familiar varieties: *modal* completion (Figure 10.3a) and *amodal* completion (Figure 10.3b). In modal completion, one experiences an illusory border between the completed shape and its surroundings—the square appears brighter than its background. In amodal completion, one has a perceptual impression of the completed shape, but no illusory border is experienced (Murray et al. 2004). In both cases, one perceptually represents the relevant object as *presently possessing* the completed shape. If we perceptually complete the missing portions of bitten shapes in either of these ways, the result would presumably be a perceptual representation of the bitten object as presently possessing its completed shape. The bitten cookie would be perceived as presently circular.

However, the 'completion' of bitten shapes differs markedly from both modal and amodal completion. Regarding the former, there seems to be no phenomenal impression of an illusory contrast border surrounding the missing portion of a bitten shape (Spröte & Fleming 2013). Regarding the latter, evidence suggests that bitten shapes are treated differently by the visual system from amodally completed shapes. Visual search data suggest that amodal completion of partially occluded objects takes place early and 'preattentively', while completion of bitten objects does

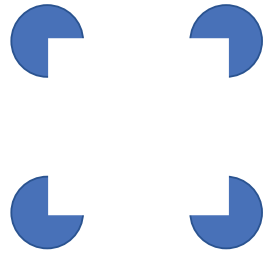


Figure 10.3a Modally completed square.

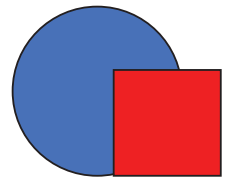


Figure 10.3b Amodally completed circle.

not (Brenner et al. 2021). Thus, the 'completion' of bitten shapes probably is not reducible to ordinary modal or amodal completion.

I recommend a different strategy. It is possible that perception might represent a past property of an object without attributing the property in the past tense *and* without representing it as a present property of the object. Instead, the property may feature as part of a complex perceptual description of a distinct present property. Call this the *complexity gambit*. In the current case, the perceptual presentist might suggest that the completed 'past' shape of the object serves as a means to representing the more complex present shape of the object, but without any explicit representation of its pastness.

Consider the object in Figure 10.4. If asked to describe its shape, it would be natural to reply: 'ellipse, save for a rectangular notch on the left'. When you answer this way, you do not commit to the claim that the object ever *was* elliptical. The appeal to ellipticality serves only as a convenient means to describing the real, complex shape of the object.

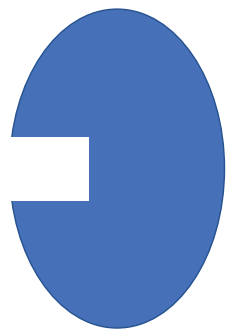


Figure 10.4 Ellipse with a notch.

There is a useful analogy to the perception of holes. Studies suggest that when we see an object with a hole, we visually represent not only the shape of the object, but also the shape of the hole it encloses (Palmer et al. 2008; Nelson et al. 2009). When asked which among several alternatives matches a target shape, participants are equally fast and accurate whether the target shape is a material surface or an immaterial hole (Nelson et al. 2014). More recently, Kim (2020) found that when participants performed the well-known 'bouba/kiki' correspondence on objects with holes, correspondence decisions were driven by the intrinsic shape of the hole, rather than the shape of its exterior.[14]

Palmer et al. (2008) propose that the visual system employs a two-part scheme to represent objects with holes. It generates one representation of the shape of the outer boundary of the surface, and another representation of the shape of the hole, with the latter marked as enclosing an empty region. Together, such representations describe the overall shape: the shape of the surface bounded by the outer contour 'minus' the shape of the hole—a 'negative part'. Because the intrinsic shape of the hole is encoded by a constituent of this representation, it is made accessible to shape recognition processes.

A similar scheme might be employed for bitten objects. Perception might complete the bitten shape by filling in the portion that has been removed, then produce a representation of the shape of the indentation caused by the bite, marked as enclosing empty space. Together, these representations would describe the overall shape: the shape of the completed counterpart minus the shape of the indentation. Critically, however, the completed shape is not represented as a *past* property of the object. The representation is silent on whether the object ever *had* the completed shape by which its present shape is described.

This story assumes that the visual system represents the intrinsic shapes of indentations caused by bites. I already mentioned such evidence in the case of holes, but there is also evidence that the shapes of other negative parts like notches or bites are perceptually represented, provided they are sufficiently salient. Mary Peterson and colleagues have shown that when we see a figure-ground display, the shape of the region seen as background, perceived as an 'intrusion' into the figural region, can be processed to the level of semantic categorization (Peterson & Skow 2008; Cacciamani et al. 2014). Cacciamani et al. (2014) showed participants figure-ground displays, such as those in Figure 10.5, in which a familiar shape was suggested on the side of the contour typically seen as background. The shapes defined by these contours nonetheless exerted semantic priming effects (faster categorization of words naming objects belonging to the same superordinate category). Thus, salient negative parts can plausibly have their shapes perceptually encoded, even when the negative part is (unlike a hole) not bounded by a closed contour.[15]

[14] See, however, Bertamini and Helmy (2012) for evidence that holes are not perceived like material objects in all respects.

[15] Note that this story only applies to *salient* negative parts. Spröte et al. (2016) found that participants plotted the symmetry axis of a completed shape, rather than the bitten shape actually shown, only when the concavity

Figure 10.5 Stimuli from Cacciamani et al. (2014). Though the white regions tend to be perceived as negative parts, their shapes are processed to a level sufficient to engage semantic priming.

Source: Cacciamani et al. (2014). Reprinted with permission of Springer Nature.

I've argued that perceptual presentists might accommodate the perception of causal history from shape through the complexity gambit. Even if we perceptually represent the 'past' shape of a distorted object, it is questionable whether we represent this shape in the past tense. Rather, it may feature as a constituent in a complex representation of the object's present shape.

At this point, one might wonder what *would* constitute good evidence for past-tensed content in perception. If the complexity gambit allows us to finesse any possible evidence against perceptual presentism, then one might worry that it is too powerful. However, I do not think this is true. In the causal history case, the complexity gambit is available only because certain present properties of the object are *naturally describable in terms of* the relevant past properties—viz., due to a close geometrical relationship between them. I suggest that progress might be made by instead investigating the representation of past properties bearing *no non-arbitrary relation* to the object's present properties. I close with a potential example of this sort.

Object Files

I argued earlier that when mnemonic representations are retained wholly offline, we can deny them perceptual status because they violate even a modest stimulus-dependence criterion. These cases involved 'pure' VWM representations completely untethered from current proximal stimulation. However, the visual system also forms complex representations containing mnemonic elements *alongside* present-tensed elements, and these representations *do* seem to satisfy stimulus-dependence.

An *object file* is a representation that sustains reference to an individual over time while storing information about that individual's current and recent features (Kahneman et al. 1992; Noles et al. 2005; Green & Quilty-Dunn 2021). Evidence for object files derives from many sources. In the object-reviewing paradigm, two

was plausibly interpreted as a bite—not when the concavities were smoothed. As Spröte et al. note, the apparently bitten concavities were more salient than the smoothed concavities due to the presence of sharp discontinuities at the opening of the concavity (Kim & Feldman 2009) and completion cues linking the contours on opposite sides of the concavity.

wireframe objects appear on a computer screen, and features (letters or shapes) appear within them before vanishing. Next, the objects move to new locations, and a new feature appears in one of them. Participants are asked whether the new feature matches either of those seen previously. Responses are faster when there is a match, but faster still when a feature reappears in the same object it appeared in initially, even though the object has shifted location. This is called an *object-specific preview benefit* (OSPB). Object files are held to underlie the OSPB. When features appear in objects at the beginning of a trial, representations of those features are entered into files dedicated to the relevant objects; these representations are retained in the files after the features vanish. Later, when a feature reappears, responses to it are facilitated if it matches information already in the relevant file.

Object files are plausibly perceptual representations. First, they subserve the vital perceptual function of apprehending *object continuity over time* (see also Munton 2022; Green 2023). Suppose that multiple objects are perceived at time T1 and time T2. The visual system faces a problem of determining, for each object at T1, which object (if any) at T2 is a continuation of it (Dawson 1991). Object-file theory claims that an object at T2 is treated as a continuation of one at T1 when a single object file targets both of them. There is evidence that the perception of object continuity in apparent motion corresponds well with object-file maintenance as indexed by the OSPB (Odic et al. 2012; although see Mitroff et al. 2005).

Object files also seem to meet the modest stimulus-dependence criterion. They are initiated in response to proximal stimulation and function to be updated in response to changes in proximal stimulation. Perhaps most obviously, an object file deployed when perceptually tracking an object must maintain a record of the object's current location, and update this record in response to signals received from the object as it moves.

Here is the threat to perceptual presentism: If (i) the mnemonic constituents of object files attribute features in the past tense, and (ii) object files—including all of their constituents—are fully perceptual, then there are some past-tensed perceptual representations, and perceptual presentism is false.

The complexity gambit does not seem applicable here because the relationship between past and present features encoded within an object file can be wholly arbitrary. If an A appears on an object and vanishes, the object need have no property afterward that is naturally describable in terms of the A. So, if object files fail to undermine perceptual presentism, it is not because they succumb to the complexity gambit.

Still, perceptual presentists might raise doubts about (i) or (ii) above.

Start with (i). I said earlier that representations in VWM plausibly mark their contents in the past tense, since this would reduce the chance of confusing objects currently before you with objects that have now vanished. But an alternative view would be that VWM representations do not *explicitly mark* their contents as past—they do not contain syntactic elements that explicitly encode this information. Instead, their functional role disposes them to be treated *as if* their contents were

past rather than present. A similar possibility arises for object files. The mnemonic constituents of an object file might play a different functional role from other constituents. This role might dispose them to be treated *as if* they attributed properties in the past tense, but without actually doing so.

One challenge for this proposal is to explain what underpins the functional-role difference between attributives for present properties and attributives for past properties, if tense is not syntactically marked. How do consuming processes 'know' to treat a given attributive in an object file as if its content were past, if this information is not explicitly encoded? Setting this question aside, however, note that functional-role properties might contribute to grounding temporal content in the absence of syntactic tense markers. If consuming processes treat a given constituent of an object file *as if* it represents the content *x was red*, this might simply *make it the case* that the constituent represents this past-tensed content, even without any discrete constituent akin to a past-tensed copula.[16]

Turning to (ii), one might suggest that object files are not *fully* perceptual representations, but *hybrid* representations with both perceptual and non-perceptual constituents. The non-mnemonic constituents are perceptual, while the mnemonic constituents are not. Thus, while object files do have past-tensed constituents, none of them are perceptual representations.

This reply prompts the question of *why* the mnemonic constituents should be dismissed as non-perceptual. I said that a vital perceptual function of object files is the apprehension of object continuity over time. One way to argue that mnemonic constituents of object files are non-perceptual would be to claim that only the non-mnemonic constituents contribute to performing this function. However, this claim is incorrect. The mnemonic constituents of object files play a key role in determining object continuity through saccades. Perception can use the information that object O *was red*, say, to determine which object visible after a saccade marks a continuation of O (Hollingworth et al. 2008; Richard et al. 2008; Schut et al. 2017).

Thus, perceptual presentists cannot withhold perceptual status from object files' mnemonic constituents on the grounds that they are not used in the performance of perceptual functions. They need another reason for doing so.

One might appeal again to stimulus-dependence. While I argued above that object files satisfy modest stimulus-dependence, perhaps not every *constituent* of an object file meets this condition. In particular, the mnemonic constituents of object files may not function to be updated in response to new sensory input.

I regard this as an empirical question. It could turn out that newly received cues to an object's past properties *are* used to update object-file representations of its past properties. If so, these representations would be stimulus-dependent

[16] Compare Block's proposal that object files possess singular content in virtue of their functional role rather than through the possession of a discrete syntactic constituent with singular content (Block 2023: ch. 5). See Green (2023) for objections to Block's model.

after all. Note also that others have objected to the view that all constituents of perceptual representations must function to be stimulus-dependent (Quilty-Dunn 2020a). So, the current strategy requires a controversially strong version of the stimulus-dependence criterion.

In any case, the current strategy raises an important, more general question. Suppose we grant that a *whole* object file is perceptual because, inter alia, it satisfies modest stimulus-dependence and subserves canonical perceptual functions. Then the question arises whether any *constituent* of a perceptual representation automatically qualifies as perceptual, 'inheriting' perceptual status from the complex of which it is a part. One view would say *yes*: the property of being a perceptual representation is like the property of being located in Massachusetts—if it is possessed by the whole, it is also possessed by the parts. Another view would say *no*: the property of being a perceptual representation is like the property of weighing ten pounds—it can be possessed by a whole but not its parts. More generally, is the property of being a perceptual representation *preserved under syntactic decomposition*?[17] If so, then object files pose a significant threat to perceptual presentism. If not, then the perceptual presentist may have an escape route.

I mention two further fallback positions for the perceptual presentist.

First, perceptual presentists might restrict their view to conscious perception, and attempt to show that the mnemonic constituents of object files—or perhaps even object files as a whole (Mitroff et al. 2005)—are always unconscious. For those primarily interested in conscious perception, I suggest that the perception of causal history from shape offers a more pressing challenge to presentism than object files.

Second, even if we grant that object files are thoroughly perceptual representations with past-tensed constituents, one could argue that there is still a *functional asymmetry* between present-tensed and past-tensed representation in perception: Whenever past-tensed representation occurs in perception, it occurs *for the purpose* of guiding production of present-tensed perceptual representations, but the converse is not true. Thus, in the case of object files, perhaps past-tensed constituents are retained solely for the purpose of guiding present-tensed representation of object continuity and persistence. If so, then perceptual presentism, strictly speaking, would be incorrect; but something would remain of the intuition that the basic function of perception is to represent present conditions. Past-tensed perception occurs only when it aids in performing this function.

I conclude that object files offer a crucial case study for perceptual presentism. They straddle the boundary between present and past. They fulfill paradigmatic perceptual functions and are modestly stimulus-dependent. However, they also possess mnemonic constituents that are vital in performing their perceptual functions.

[17] *Decomposition* is the key direction here. One might hold that a non-perceptual representation could have a perceptual representation as a constituent, but that no perceptual representation could have a non-perceptual representation as a constituent.

6. Conclusion

Perceptual presentists hold that perception is restricted to representing present conditions. While perceptual presentism is commonly assumed, it is rarely defended. Moreover, various strands of evidence suggest that perception is attuned to the past in lawlike ways. This chapter has distinguished three forms of perceptual sensitivity to the past and considered their bearing on perceptual presentism. I argued that only the third form constitutes a genuine threat to the view. While the case for this form of sensitivity remains inconclusive, I suggest the most compelling challenges to perceptual presentism derive from complex representations that integrate mnemonic and present-tensed elements in performing canonical perceptual functions.[18]

References

Adam, K. C. S., Rademaker, R. L., & Serences, J. T. (2021). Evidence for, and challenges to, sensory recruitment models of visual working memory. In T. F. Brady & W. A. Bainbridge (eds.), *Visual Memory* (pp. 5–26). New York: Routledge.

Anstis, S., Verstraten, F. A., & Mather, G. (1998). The motion after effect. *Trends in Cognitive Sciences*, 2(3), 111–17.

Balcetis, E., & Dunning, D. (2010). Wishful seeing: More desired objects are seen as closer. *Psychological Science*, 21(1), 147–52.

Bao, M., Yang, L., Rios, C., He, B., & Engel, S. A. (2010). Perceptual learning increases the strength of the earliest signals in visual cortex. *Journal of Neuroscience*, 30(45), 15080–84.

Barkasi, M. (2021). Memory as sensory modality, perception as experience of the past. *Review of Philosophy and Psychology*. DOI: 10.1007/s13164-021-00598-7.

Beck, J. (2018). Marking the perception–cognition boundary: The criterion of stimulus-dependence. *Australasian Journal of Philosophy*, 96(2), 319–34.

Bertamini, M., & Helmy, M. S. (2012). The shape of a hole and that of the surface-with-hole cannot be analyzed separately. *Psychonomic Bulletin & Review*, 19(4), 608–16.

Block, N. (2018). If perception is probabilistic, why does it not seem probabilistic? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1755). 20170341.

Block, N. (2023). *The Border between Seeing and Thinking*. Oxford: Oxford University Press.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision*, 11(5):4, 1–34.

[18] Thanks to Alex Byrne, Jessie Munton, and Ian Phillips for detailed comments on earlier drafts of this chapter. Thanks also to audiences at the *Space, Time, and Memory* workshop in Tucson, the University of Glasgow, and the WashU Mind and Perception Group for helpful feedback. I have also benefited from conversations about these issues with Eric Mandelbaum, Casey O'Callaghan, Shen Pan, Mary Peterson, and Jake Quilty-Dunn.

Brenner, E., Hurtado, S. S., Arias, E. A., Smeets, J. B., & Fleming, R. W. (2021). Searching for strangely shaped cookies: Is taking a bite out of a cookie similar to occluding part of it? *Perception*, 50(2), 140–53.

Burr, D., & Ross, J. (2008). A visual sense of number. *Current Biology*, 18(6), 425–28.

Butler, A., Oruc, I., Fox, C. J., & Barton, J. J. (2008). Factors contributing to the adaptation after effects of facial expression. *Brain Research*, 1191, 116–26.

Byrne, A. (2018). *Transparency and Self-Knowledge*. Oxford: Oxford University Press.

Cacciamani, L., Mojica, A. J., Sanguinetti, J. L., & Peterson, M. A. (2014). Semantic access occurs outside of awareness for the ground side of a figure. *Attention, Perception, & Psychophysics*, 76(8), 2531–47.

Carey, S. (2009). *The Origin of Concepts*. Oxford: Oxford University Press.

Ceylan, G., Herzog, M. H., & Pascucci, D. (2021). Serial dependence does not originate from low-level visual processing. *Cognition*, 212, 104709.

Chaisilprungraung, T., German, J., & McCloskey, M. (2019). How are object shape axes defined? Evidence from mirror-image confusions. *Journal of Experimental Psychology: Human Perception and Performance*, 45(1), 111–24.

Chen, Y. C., & Scholl, B. J. (2016). The perception of history: Seeing causal history in static shapes induces illusory motion perception. *Psychological Science*, 27(6), 923–30.

Chuard, P. (2011). Temporal experiences and their parts. *Philosophers' Imprint*, 11(11), 1–28.

Clark, A. (1992). Presence of a symbol. *Connection Science*, 4(3–4), 193–205.

Clark, A. (2015). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford: Oxford University Press.

Collins, T. (2019). The perceptual continuity field is retinotopic. *Scientific Reports*, 9(1), 1–6.

Connolly, K. (2019). *Perceptual Learning: The Flexibility of the Senses*. Oxford: Oxford University Press.

Connor, A., & Smith, J. (2019). The perceptual present. *The Philosophical Quarterly*, 69(277), 817–37.

Czoschke, S., Fischer, C., Beitner, J., Kaiser, J., & Bledowski, C. (2019). Two types of serial dependence in visual working memory. *British Journal of Psychology*, 110(2), 256–67.

Dainton, B. (2022). Temporal consciousness. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/sum2022/entries/consciousness-temporal/&gt.

Dawson, M. R. (1991). The how and why of what went where in apparent motion: Modeling solutions to the motion correspondence problem. *Psychological Review*, 98(4), 569–603.

Dosher, B., & Lu, Z. L. (2017). Visual perceptual learning and models. *Annual Review of Vision Science*, 3, 343–63.

Durgin, F. H., DeWald, D., Lechich, S., Li, Z., & Ontiveros, Z. (2011). Action and motivation:Measuring perception or strategies? *Psychonomic Bulletin & Review*, 18(6), 1077–82.

Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for 'top-down' effects. *Behavioral and Brain Sciences*, 39, 1–19.

Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience*, *17*(5), 738–43.

Flombaum, J. I., Scholl, B. J., & Pylyshyn, Z. W. (2008). Attentional resources in visual tracking through occlusion: The high-beams effect. *Cognition*, 107(3), 904–31.

Fritsche, M., & de Lange, F. P. (2019). The role of feature-based attention in visual serial dependence. *Journal of Vision*, 19(13), 21, 1–13.

Fritsche, M., Mostert, P., & de Lange, F. P. (2017). Opposite effects of recent history on perception and decision. *Current Biology*, 27(4), 590–95.

Furmanski, C. S., Schluppeck, D., & Engel, S. A. (2004). Learning strengthens the response of primary visual cortex to simple patterns. *Current Biology*, 14(7), 573–78.

Gale, R. (ed.). (1968). *The Philosophy of Time*. Sussex: Harvester.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, 49(1), 585–612.

Green, E. J. (2020). The perception-cognition border: A case for architectural division. *Philosophical Review*, 129(3), 323–93.

Green, E. J. (2023). The perception-cognition border: Architecture or format? In B. McLaughlin & J. Cohen (eds.), *Contemporary Debates in Philosophy of Mind* (pp. 469–93). Oxford: Blackwell.

Green, E. J., & Quilty-Dunn, J. (2021). What is an object file? *The British Journal for the Philosophy of Science*, 72(3), 665–99.

Gross, S. (2017). Cognitive penetration and attention. *Frontiers in Psychology*, 8, 221.

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, 458(7238), 632–35.

Hein, E., Stepper, M. Y., Hollingworth, A., & Moore, C. M. (2021). Visual working memory content influences correspondence processes. *Journal of Experimental Psychology: Human Perception and Performance*, 47(3), 331–43.

Herzog, M. H., Drissi-Daoudi, L., & Doerig, A. (2020). All in good time: Long-lasting postdictive effects reveal discrete perception. *Trends in Cognitive Sciences*, 24(10), 826–37.

Hoerl, C. (2018). Experience and time: Transparency and presence. *Ergo*, 5(5), 127–51.

Hogendoorn, H. (2020). Motion extrapolation in visual processing: Lessons from 25 years of flash-lag debate. *Journal of Neuroscience*, 40(30), 5698–705.

Hollingworth, A., Richard, A. M., & Luck, S. J. (2008). Understanding the function of visual short-term memory: Transsaccadic memory, object correspondence, and gaze correction. *Journal of Experimental Psychology: General*, 137(1), 163–81.

Hubbard, T. L., Hutchison, J. L., & Courtney, J. R. (2010). Boundary extension: Findings and theories. *Quarterly Journal of Experimental Psychology*, 63(8), 1467–94.

Humphreys, G. W., & Quinlan, P. T. (1988). Priming effects between two-dimensional shapes. *Journal of Experimental Psychology: Human Perception and Performance*, 14(2), 203–20.

Icard, T. (2016). Subjective probability as sampling propensity. *Review of Philosophy and Psychology*, 7(4), 863–903.

Intraub, H., & Dickinson, C. A. (2008). False memory 1/20th of a second later: What the early onset of boundary extension reveals about perception. *Psychological Science*, 19(10), 1007–14.

James, W. (1890). *The Principles of Psychology*. New York: Dover.

Jehee, J. F., Ling, S., Swisher, J. D., van Bergen, R. S., & Tong, F. (2012). Perceptual learning selectively refines orientation representations in early visual cortex. *Journal of Neuroscience*, 32(47), 16747–53.

Jones, P. D., & Holding, D. H. (1975). Extremely long-term persistence of the McCollough effect. *Journal of Experimental Psychology: Human Perception and Performance*, 1(4), 323–27.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, 24(2), 175–219.

Kang, M. S., Hong, S. W., Blake, R., & Woodman, G. F. (2011). Visual working memory contaminates perception. *Psychonomic Bulletin & Review*, 18(5), 860–69.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, 55, 271–304.

Kim, S. H. (2020). Bouba and Kiki inside objects: Sound-shape correspondence for objects with a hole. *Cognition*, 195, 104132.

Kim, S. H., & Feldman, J. (2009). Globally inconsistent figure/ground relations induced by a negative part. *Journal of Vision*, 9(10), 8, 1–13.

Kominsky, J. F., & Scholl, B. J. (2020). Retinotopic adaptation reveals distinct categories of causal perception. *Cognition*, 203, 104339.

Kriegel, U. (2015). Experiencing the present. *Analysis*, 75(3), 407–13.

Landman, R., Spekreijse, H., & Lamme, V. A. (2003). Large capacity storage of integrated objects before change blindness. *Vision Research*, 43(2), 149–64.

Law, C. T., & Gold, J. I. (2008). Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area. *Nature Neuroscience*, 11(4), 505–13.

Leyton, M. (1989). Inferring causal history from shape. *Cognitive Science*, 13(3), 357–87.

Leyton, M. (1992). *Symmetry, Causality, Mind*. Cambridge, MA: MIT Press.

Macpherson, F. (2012). Cognitive penetration of colour experience: Rethinking the issue in light of an indirect mechanism. *Philosophy and Phenomenological Research*, 84(1), 24–62.

Mandelbaum, E. (2018). Seeing and conceptualizing: Modularity and the shallow contents of perception. *Philosophy and Phenomenological Research*, 97(2), 267–83.

Mitroff, S. R., Scholl, B. J., & Wynn, K. (2005). The relationship between object files and conscious perception. *Cognition*, 96(1), 67–92.

Munton, J. (2022). How to see invisible objects. *Noûs*, 56(2), 343–65.

Murray, M. M., Foxe, D. M., Javitt, D. C., & Foxe, J. J. (2004). Setting boundaries: Brain dynamics of modal and amodal illusory shape completion in humans. *Journal of Neuroscience*, 24(31), 6898–903.

Nanay, B. (2021). Boundary extension as mental imagery. *Analysis*, 81(4), 647–56.

Nelson, R., Reiss, J. E., Gong, X., Conklin, S., Parker, L., & Palmer, S. E. (2014). The shape of a hole is perceived as the shape of its interior. *Perception*, 43(10), 1033–48.

Nelson, R., Thierman, J., & Palmer, S. E. (2009). Shape memory for intrinsic versus accidental holes. *Perception & Psychophysics*, 71(1), 200–06.

Noles, N. S., Scholl, B. J., & Mitroff, S. R. (2005). The persistence of object file representations. *Perception & Psychophysics*, 67(2), 324–34.

Odic, D., Roth, O., & Flombaum, J. I. (2012). The relationship between apparent motion and object files. *Visual Cognition*, 20(9), 1052–81.

Orlandi, N. (2016). Bayesian perception is ecological perception. *Philosophical Topics*, 44(2), 327–52.

Palmer, S., Davis, J., Nelson, R., & Rock, I. (2008). Figure–ground effects on shape memory for objects versus holes. *Perception*, 37(10), 1569–86.

Pascucci, D., Mancuso, G., Santandrea, E., Della Libera, C., Plomp, G., & Chelazzi, L. (2019). Laws of concatenated perception: Vision goes for novelty, decisions for perseverance. *PLoS Biology*, 17(3), e3000144.

Peacocke, C. (1999). *Being Known*. Oxford: Oxford University Press.

Peterson, M. A., & Skow, E. (2008). Inhibitory competition between shape properties in figure-ground perception. *Journal of Experimental Psychology: Human Perception and Performance*, 34(2), 251–67.

Phillips, B. (2019). The shifting border between perception and cognition. *Noûs*, 53(2), 316–46.

Phillips, I. B. (2011). Perception and iconic memory: What Sperling doesn't show. *Mind & Language*, 26(4), 381–411.

Phillips, I. B. (2014). Experience of and in time. *Philosophy Compass*, 9(2), 131–44.

Phillips, I. B., & Firestone, C. (forthcoming). Visual adaptation and the purpose of perception Analysis.

Pourtois, G., Rauss, K. S., Vuilleumier, P., & Schwartz, S. (2008). Effects of perceptual learning on primary visual cortex activity in humans. *Vision Research*, 48(1), 55–62.

Pryor, J. (2000). The skeptic and the dogmatist. *Noûs*, 34(4), 517–49.

Pylyshyn, Z. W. (2003). *Seeing and Visualizing: It's Not What You Think*. Cambridge, MA: MIT Press.

Quilty-Dunn, J. (2020a). Concepts and predication from perception to cognition. *Philosophical Issues*, 30(1), 273–92.

Quilty-Dunn, J. (2020b). Attention and encapsulation. *Mind & Language*, 35(3), 335–49.

Quinlan, P. T., & Humphreys, G. W. (1993). Perceptual frames of reference and two-dimensional shape recognition: Further examination of internal axes. *Perception*, 22(11), 1343–64.

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

Reid, T. (1785/2011). *Essays on the Intellectual Powers of Man*. Cambridge: Cambridge University Press.

Rescorla, M. (2020). A realist perspective on Bayesian cognitive science. In T. Chan & A. Nes (eds.), *Inference and Consciousness* (pp. 40–73). New York: Routledge.

Richard, A. M., Luck, S. J., & Hollingworth, A. (2008). Establishing object correspondence across eye movements: Flexible use of spatiotemporal and surface feature information. *Cognition*, 109(1), 66–88.

Rock, I. (1982). Inference in perception. In P. D. Asquith & T. Nickles (eds.), *Proceedings of the Biennial Meeting of the Philosophy of Science Association, volume 2: Symposia and Invited Papers* (pp. 525–40). East Lansing, MI: Philosophy of Science Association.

Rolfs, M., Dambacher, M., & Cavanagh, P. (2013). Visual adaptation of the perception of causality. *Current Biology*, 23(3), 250–54.

Sanborn, A. N., & Chater, N. (2016). Bayesian brains without probabilities. *Trends in Cognitive Sciences*, 20(12), 883–93.

Schmidt, F., & Fleming, R. W. (2016). Visual perception of complex shape-transforming processes. *Cognitive Psychology*, 90, 48–70.

Schmidt, F., & Fleming, R. W. (2018). Identifying shape transformations from photographs of real objects. *PloS One*, 13(8), e0202115.

Schmidt, F., Phillips, F., & Fleming, R. W. (2019). Visual perception of shape-transforming processes: 'Shape scission'. *Cognition*, 189, 167–80

Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology*, 38(2), 259–90.

Schoups, A., Vogels, R., Qian, N., & Orban, G. (2001). Practising orientation identification improves orientation coding in V1 neurons. *Nature*, 412(6846), 549–53.

Schut, M. J., Fabius, J. H., Van der Stoep, N., & Van der Stigchel, S. (2017). Object files across eye movements: Previous fixations affect the latencies of corrective saccades. *Attention, Perception, & Psychophysics*, 79(1), 138–53.

Scocchia, L., Valsecchi, M., Gegenfurtner, K. R., & Triesch, J. (2013). Visual working memory contents bias ambiguous structure from motion perception. *PloS One*, 8(3), e59217.

Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science*, 20(2), 207–14.

Shea, N. (2015). Distinguishing top-down from bottom-up effects. In D. Stokes, M. Matthen, & S. Biggs (eds.), *Perception and Its Modalities* (pp. 73–91). Oxford: Oxford University Press.

Silins, N. (2021). Perceptual experience and perceptual justification. In E. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/win2021/entries/perception-justification/.

Skow, B. (2011). Experience and the passage of time. *Philosophical Perspectives*, 25, 359–87.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, 74(11), 1–29.

Spröte, P., & Fleming, R. W. (2013). Concavities, negative parts, and the perception that shapes are complete. *Journal of Vision*, 13(14), 3, 1–23.

Spröte, P., & Fleming, R. W. (2016). Bent out of shape: The visual inference of non-rigid shape transformations applied to objects. *Vision Research*, 126, 330–46.

Spröte, P., Schmidt, F., & Fleming, R. W. (2016). Visual perception of shape altered by inferred causal history. *Scientific Reports*, 6(1), 1–11.

Suchow, J. W., Fougnie, D., Brady, T. F., & Alvarez, G. A. (2014). Terms of the debate on the format and structure of visual memory. *Attention, Perception, & Psychophysics*, 76(7), 2071–79.

Suddendorf, T., & Corballis, M. C. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans? *Behavioral and Brain Sciences*, 30(3), 299–313.

Teng, C., & Kravitz, D. J. (2019). Visual working memory directly alters perception. *Nature Human Behaviour*, 3(8), 827–36.

Thompson, P., & Burr, D. (2009). Visual after effects. *Current Biology*, 19(1), R11–R14.

Viera, G. (2022). The perceived unity of time. *Mind & Language*, 37(4), 638–58.

Webster, M. A. (2011). Adaptation and visual coding. *Journal of Vision*, 11(5), 3, 1–23.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, 5(6), 598–604.

White, P. A. (2018). Is the perceived present a predictive model of the objective present? *Visual Cognition*, 26(8), 624–54.

White, P. A. (2020). The perceived present: What is it, and what is it there for? *Psychonomic Bulletin & Review*, 27(4), 583–601.

Yan, Y., Rasch, M. J., Chen, M., Xiang, X., Huang, M., Wu, S., & Li, W. (2014). Perceptual training continuously refines neuronal population codes in primary visual cortex. *Nature Neuroscience*, 17(10), 1380–87.

11

Experience Replay Algorithms and the Function of Episodic Memory

Alexandria Boyle

Introduction

Think back to when you woke up this morning.[1] Maybe you snoozed your alarm several times, or perhaps you jumped out of bed straight away, ready to face the day. Maybe you had a leisurely breakfast and chatted to your family, or maybe you had to tumble out of the house in a hurry with nothing but a coffee to go. Whatever happened, as you think back to this morning you may have the feeling that you're 'replaying' your experiences in your mind's eye. You might mentally smell the coffee, see it foam or hear the hiss as it comes out of the machine. This phenomenon—experience replay—is characteristically associated with *episodic memory*: memory for personally experienced past events.

Episodic memory presents several puzzles for philosophers and scientists of memory. Among them is the question of its function: what does episodic memory contribute to the cognitive system (Cummins 1975)? To put it another way, what is it that we're able to do because we can remember and 'replay' past events? What makes this puzzle challenging to resolve is the difficulty of evaluating competing theories of episodic memory's function. To do so, we'd ideally need an agent which had episodic memory, whose episodic memory capacity we could 'switch off' whilst leaving its other cognitive functions intact, so that we could isolate the unique contributions episodic memory makes. For reasons I'll outline, it is not easy to find an agent like this in the biological realm. But recent developments in AI suggest an intriguing possibility: perhaps we could make one.

In recent years, a number of algorithms have been developed which exploit an 'experience replay' mechanism, drawing direct inspiration from episodic memory in biological systems (Hassabis et al. 2017). These algorithms record details of their 'experiences' and replay these experiences after the fact. The replay mechanism is

[1] Earlier versions of this chapter were presented to audiences at the London School of Economics, Kings College London, the London Mind Group, Duke University, and the University of Arizona. I'm grateful to all those present for helpful feedback and discussion. Thanks also to Andrea Blomkvist, Julia Haas, and an anonymous reviewer for comments on an earlier draft. This research was supported by a UKRI Future Leaders Fellowship (grant number MR/W00741X/1). There are no data associated with this work.

Alexandria Boyle, *Experience Replay Algorithms and the Function of Episodic Memory*. In: *Space, Time, and Memory*. Edited by: Lynn Nadel and Sara Aronowitz, Oxford University Press.
DOI: 10.1093/oso/9780192882547.003.0011

often separable from the rest of the cognitive architecture, making it possible to directly assess its contributions to the system's capabilities. In this chapter, my question is whether, by investigating the effects of experience replay on artificial agents, we stand to learn about episodic memory's function in biological systems. For this to be a promising research strategy, there would need to be more than a superficial resemblance between biological and artificial experience replay. The two would need to be meaningfully similar, such that one could hope to draw justified (if defeasible) inductive inferences about biological episodic memory from artificial experience replay. Whilst there are significant differences between these phenomena, my argument here will be that the two are sufficiently similar in relevant ways to ground such inferences—so, by investigating these algorithms we stand to gain defeasible evidence about the function of episodic memory.

I begin in §2 by describing DeepMind's deep-Q network (DQN) algorithm (Mnih et al. 2015), the experience replay algorithm I take as my case study in this chapter.[2] In §3, I argue that in order to inform our accounts of episodic memory's function, DQN would need to resemble episodic memory at the algorithmic level. I show that there are significant similarities between the two at the algorithmic level: both exploit detailed, iconic representations integrating multidimensional information about specific past events. In §4, I apply this result to the debate about episodic memory's function. After briefly summarizing the debate between simulationist and mnemonic accounts of episodic memory's function, I argue that DQN provides some support to the mnemonic view. I also acknowledge several differences between DQN and episodic memory. I argue that for some purposes, these differences will not matter: DQN can fruitfully serve as an idealized model of episodic memory whilst differing from it in various ways. In other contexts, however, these differences will matter. This reveals ways in which we might look to extend or adapt experience replay algorithms to more closely resemble episodic memory in the relevant respects. Doing so would facilitate the evaluation of more granular theories about episodic memory's operations and contributions to cognition. §5 concludes.

The DQN Algorithm

The Arcade Learning Environment is a platform which uses Atari 2600 games, like Breakout, Pong, and Space Invaders, to evaluate AI systems (Bellemare et al. 2013). In 2015, DeepMind reported that their DQN algorithm had achieved a new state of the art in the Arcade Learning Environment, achieving better scores than the previous best-performing algorithm in forty-three of forty-nine games tested (Mnih et al. 2015). DQN's performance was comparable to that of a professional human games tester, achieving at least 75% of the human's score on more than half of the games

[2] Given the differences between DQN and other implementations of experience replay, the conclusions I draw here can't be generalized to other architectures without argument. I discuss some other experience replay algorithms briefly in §4.

sampled. Key to DQN's success was a 'biologically inspired' experience replay mechanism. Unlike previous systems, which learned from each 'experience' and moved on, this algorithm recorded its experiences and replayed them after the fact.

DQN is a reinforcement-learning algorithm. Reinforcement learning is a machine learning paradigm in which an agent learns by taking actions in its environment. Reinforcement-learning problems are commonly modelled as Markov decision processes. In a Markov decision process, an agent interacts with its environment at several discrete time stems, t = 1, 2, ..., n. At each time-step, the agent observes the state of the environment (S_t), selects an action (A_t), receives a reward (R_t) and observes the resulting state of the environment at the next time-step (S_{t+1}) (Sutton & Barto 2018). In the Arcade Learning Environment, the reward is the change—positive or negative—in game score. Over time, the idea is that the agent will learn a policy which maximizes its future rewards. Roughly speaking, a policy is a state–action mapping; something that determines how the agent will act in any given state. The agent may initially act randomly, but by observing the effects and rewards produced by its actions, it develops more sophisticated policies, enabling it to respond effectively to the environment (Figure 11.1).

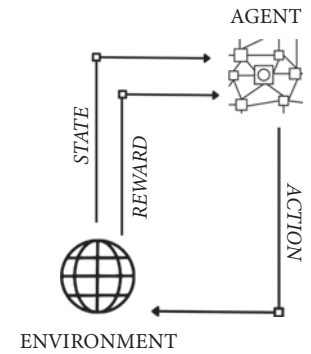


Figure 11.1 Agent–environment interactions in a Markov decision process.

DQN implements a form of reinforcement learning known as Q-learning. Q-learning involves learning the expected reward (the 'Q-value') of the actions in each scenario. In 'vanilla' Q-learning (i.e., the most basic form), Q-values are represented in a look-up table mapping state/action pairs to Q-values (Figure 11.2). When the agent acts, the Q-table is updated to reflect new information about the relevant state/action pairs. DQN is a variation on Q-learning in which the Q-table is replaced by a neural network mapping input states (i.e., observed states of the environment) to action/Q-value pairs.

At each time-step, the agent selects an action using an 'ε-greedy' policy: it selects a random action with probability ε, and selects the action predicted to have the highest value with probability 1-ε. The value of ε is adjusted during training such that the agent begins by randomly exploring the available actions, later shifting towards exploiting what it has learned. Whenever the agent takes an action, the

STATE → ACTION↓	S1	S2	S3
A1	50	78	4
A2	67	9	33
A3	1	45	70

Figure 11.2 Q-table.

divergence between the expected and actual reward is used as an error signal to train the Q-learning network.

The important thing about DQN for our purposes is its use of experience replay. In experience replay, the agent's experience at each time-step is recorded. An experience consists of a 4-tuple representation (S_t, A_t, R_t, S_{t+1}), where:

- S_t = the state of the environment at t
- A_t = the action taken at t
- R_t = the reward obtained at t
- S_{t+1} = the state of the environment at $t+1$.

These 4-tuple experiences are pooled in a store called the 'episodic buffer'. The oldest experiences are deleted when the episodic buffer's finite capacity is reached.

Importantly for our purposes, the state representations (S_t) are more complex than this brief sketch suggests. 'S_t' in fact represents a sequence of observations (X_t) and actions (A_t) over the m time-steps leading up to and including t. The number of time-steps per sequence (m) is variable; in DeepMind's original implementation, m = 4. S_t is then shorthand for the sequence: X_{t-3}, A_{t-3}, X_{t-2}, A_{t-2}, X_{t-1}, A_{t-1}, X_t. Each X is an observation of the screen of the Atari emulator, represented as pixel vector. Simplifying, this is generated by taking a screenshot from the Atari system, applying some minimal pre-processing to remove artifacts, standardizing the size of the screenshot and converting it to greyscale. Each pixel in the resulting 84 × 84 pixel greyscale image can now be represented as a single number, the luminance value corresponding to its particular shade of grey. These luminance values are used to convert the image into a two-dimensional numerical array. The two dimensions of the array correspond to image height and width; each number in the array picks out the pixel in the corresponding image location and represents its luminance. The sequence, S_t, is then a temporally ordered sequence of pixel vectors interspersed with the actions taken at the corresponding time-step.

The Q-learning network learns 'off-policy', by training itself on minibatches of experiences selected at random from the episodic buffer, rather than directly on the agent's most recent experience (see Figure 11.3). One advantage of using past experiences for learning in this way is that each experience can be used for learning several times. Another is that it eliminates the correlations which

would otherwise hold between consecutive training samples, which are both inefficient for learning and increase the chances of the system getting stuck in local minima—that is, becoming committed to suboptimal strategies. Consistently with experience replay conferring these learning advantages, disabling replay significantly worsens the network's performance, sometimes by an order of magnitude (Mnih et al. 2015: Table 3).

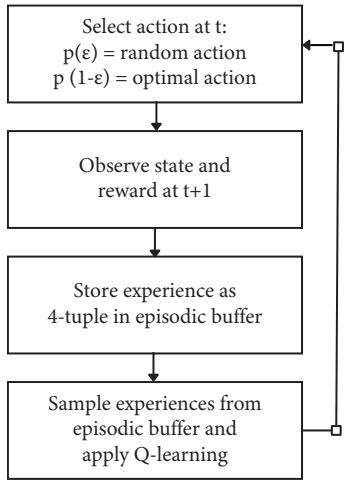


Figure 11.3 Simplified sketch of the DQN algorithm.

Representations in Episodic Memory and DQN

As noted in §1, DQN's reliance on experience replay is suggestive because of the characteristic association between experience replay and episodic memory in humans. But to determine whether this is any more than suggestive, we'd need to know whether this is a *meaningful* similarity between episodic memory and DQN.

In general, cognitive systems can be described and explained at various 'levels. David Marr's (1982) prominent account distinguishes the computational, algorithmic, and implementation levels. The computational level is the level at which we describe *what* the system is doing and *why*; the algorithmic level describes what representations are used and what algorithms are used to process them; the implementation level describes the physical structures realizing this processing. There is some independence between these levels, meaning that two systems might resemble one another at one level whilst being very different at another.

In the current context, we are ultimately interested in understanding the cognitive role function of episodic memory: what it does, and how, by doing that, it contributes to the capacities of the cognitive systems of which it's a part (Cummins 1975). What makes this puzzling is that, on an intuitive picture of what episodic memory does, it's not obvious that it has a distinctive contribution to make. Its central job is to carry detailed information about specific past events. But we have a

general-purpose memory system in the form of semantic memory—a decontextualized memory store carrying information about words, conceptual relations or facts. So, we might wonder, what is the point of having *another* memory system dedicated to the storage, encoding and retrieval of detailed information about specific past events when, after all, much of this information is surely redundant, and the specific events will never come around again?[3] Put in these terms we can see that our question is pitched at the algorithmic level: what is the point of having a memory system that uses *these* kinds of representations in this way? For DQN to aid us in answering this question, we would therefore need to establish that it resembled episodic memory at the algorithmic level.[4] The task of this section is to establish that there are reasons for thinking this is true, by showing that DQN and episodic memory exploit similar kinds of representations.

The first step is to give an account of the representations exploited by biological episodic memory. When Endel Tulving introduced the distinction between episodic and semantic memory (Tulving 1972), he distinguished the two in terms of their content. Episodic memory was a matter of remembering *what* happened, *where*, and *when*. However, Tulving and many others came to see this characterization of episodic memory as inadequate, for two reasons. First, there are examples of episodic memories in which one or more of the 'what-where-when' components is missing. Second, it is not uncommon to semantically remember the 'what-where-when' of events, including events which one could not episodically remember, such as the Battle of Hastings. These seem significant problems: ideally, an account of episodic memory ought at least to capture what *in general* distinguishes it from its most salient contrast class (Boyle 2020b).

In response to this, Tulving and others came to place greater weight on the characteristic *phenomenology* of episodic recollection: the experience described in terms of 'replaying' past events (Tulving 2005).[5] To place too much weight on *experience* in characterizing episodic memory feels problematic, however. Most significantly for our purposes, this provides little guidance when it comes to evaluating the significance of algorithms like DQN. Since, plausibly, no current AI systems are conscious, no current architecture will exhibit the relevant phenomenology, suggesting that no current experience replay architecture implements episodic memory. To be clear, that may well be the right result. But this does not settle the question of whether architectures like DQN can inform our understanding of biological episodic memory, since they might nevertheless be similar in significant respects. Nevertheless, as I've argued elsewhere (Boyle 2020b), characterizing episodic memory in terms of experience replay can be fruitful, since careful reflection on what is involved in 'replaying' past events allows us to characterize the *representations* involved in

[3] See Brown (2024: sec. 3) for a compelling discussion of this puzzle.

[4] This means that we need not be troubled in this context by the obvious fact that DQN and biological episodic memory processes differ at the level of implementation: DQN is implemented in ordinary computer hardware, and episodic memory in brain structures and processes.

[5] Following Tulving (2005), this experience is sometimes described in terms of 'autonoetic consciousness' and 'chronesthesia'. I find this terminology unhelpful for reasons I've outlined elsewhere (Boyle 2020a), and will avoid it here.

episodic memory in more granular terms than 'what-where-when'. This reveals an account of episodic memory which both distinguishes it from semantic memory and facilitates comparisons with experience replay algorithms: we can determine whether experience replay algorithms involve representations which carry similar information and are structured in similar ways.

To say that episodic memory characteristically involves experience replay involves taking on some substantial commitments about the content episodic memory carries. First, it suggests that episodic memory carries detailed spatial information about an event. This is not merely to say that it carries information about *where* the event occurred. In fact, one might episodically remember an event without being able to pinpoint its location. But episodically remembering an event typically involves remembering what we might call the event's 'internal' spatial features, that is, the spatial context in which the event occurred. For instance, if you remember having breakfast this morning, you might remember the room you ate in, how the furniture was arranged in that room, where your cereal bowl was relative to the table, where the cereal was relative to the bowl, and so on. That episodic memory captures this kind of contextual spatial information about an event is central to accounts characterizing episodic memory in terms of 'scene construction' (Boyle 2020a; Clayton & Russell 2009; Rubin & Umanath 2015).

In addition, to the extent that episodic memory prototypically involves replaying events, it must carry similar information about the event's temporal features. Once again, this is not a matter of remembering *when* the event occurred. In fact, episodic memory need not involve representing events as past at all (Boyle 2020a). But episodically remembering an event, such that one could in principle mentally replay or re-experience it, must involve remembering its 'internal' temporal features, that is, the order in which its component parts occurred (Boyle 2020b). So, again, if you remember having breakfast this morning, you might remember that your cereal bowl was full at the beginning and became progressively emptier, that the cereal was crispy at the beginning but became progressively soggier, and so on.

Spatial and temporal information are of course not the only kinds of information episodic memory characteristically carries about an event. Given that episodic memories are of events you personally witnessed or were involved in, we might add that episodic memory can carry self-referential information, such as information about how you were involved, what you were thinking, feeling, or perceiving (Boyle 2020b).

So, we now have a somewhat fuller characterization of the sort of *content* episodic memory prototypically carries, namely, detailed, multidimensional information about an event, including the spatial and temporal organization of the event, and the remembering subject's involvement in, perception of and thoughts and feelings about the event at the time of its occurrence. This is by no means a novel view of episodic memory's content: a number of recent accounts emphasize the fact that episodic memory carries event information across a number of quality dimensions

(Brown 2024; Gershman & Daw 2017). This picture of episodic memory's content is also reflected in the methodologies used to investigate episodic memory in cognitive science. For instance, the Autobiographical Interview Questionnaire distinguishes the episodic and semantic aspects of an autobiographical memory by coding reported details as either 'internal' (episodic) or 'external' (semantic), where 'internal details' are comments on the event's spatiotemporal structure, and the subject's thoughts, emotions, or perceptions during the event (Levine et al. 2002). Several methods for detecting episodic memory in animals also rest on the idea that episodic memory carries detailed multidimensional event information. For instance, the 'what-where-which' protocol investigates whether an animal can discriminate between similar events which occurred in distinct spatial contexts (Eacott et al. 2005), whilst 'source monitoring' studies investigate whether an animal remembers contextual detail about an event besides what happened, where, and when (Crystal et al. 2013).

Whilst 'replaying' an event involves recalling detailed multidimensional event information, the various details are not recalled separately from one another. Rather, these various details are presented to us as a package: they seem to be integrated into a single, structured representation of the event as a whole (Boyle 2021; Rubin & Umanath 2015). Once again, this idea is reflected in methods used to detect episodic memory. The 'integration' criterion for detecting episodic memory in animals investigates whether the various datapoints an animal remembers about an event are integrated into a single, unified representation (Clayton et al. 2001). One operationalization of this is that memory is integrated when retrieval of one piece of information encoded in the memory predicts retrieval of the rest (Clayton et al. 2003). Another is that a memory is integrated if it is resistant to interference from memories for events composed out of similar informational components—that is, if the subject can remember and distinguish similar but distinct events (Crystal & Smith 2014). Underlying the idea that episodic memories should be integrated is the idea that episodic memories combine diverse datapoints into a single representational unit.

It is tempting to cash this out as a claim about episodic memory's format: perhaps episodic memory has a (partly) iconic format. Iconic formats are characterized by structural isomorphism. That is to say, the structure of the representation mirrors or 'maps on to' the structure of the thing being represented, and that mapping is semantically significant (Lee et al. 2022; Shea 2014). Iconic representations are typically informationally rich, integrating detailed, often multidimensional information, into a single representational unit. So, we might think a plausible hypothesis about experience replay is that it involves representations with an iconic format, and this explains why they integrate rich multidimensional information about events in a way that is resistant to interference from similar memories. One thing that renders this intuitively plausible is the way experience replay seems to represent the temporal properties of remembered events. Replayed experiences seem to 'unfold' over time in the mind's eye, in a way that mirrors the represented unfolding of the event

(Boyle 2020b). Here, the temporal features of the representation appear to map onto the temporal features of the represented event.

For these kinds of reasons, Nikola Andonovski (2022) argues that episodic memories are 'structure-preserving models' of past events. On this view, episodic memories and other forms of episodic representation are abstract mental models which mirror the spatiotemporal structure of represented events, and perhaps their structure across other quality dimensions. This mirroring is unlikely to be a strict isomorphism; it may be approximate or simplified. Importantly, this structure-preserving format is unlikely to exhaust the representational features of episodic memory: given the pervasive interactions between episodic and semantic memory (Aronowitz 2022; Boyle 2021), instances of episodic remembering will almost always involve conceptual or semantic elements in addition to the structure-preserving model at their core.

On the view of episodic memory we've arrived at, then, it involves retrieving representations which carry detailed multidimensional information about a remembered event. Minimally, this includes its spatial and temporal organization, perhaps along with information about the event's other qualities, the subject's background knowledge, and the subject's involvement in, perception of, and thoughts and feelings about the event. These details seem to be retrieved as a package, indicating that the representations involved unite these details into an integrated representational whole. This suggests the representations involved may be at least partly iconic, perhaps taking the form of structure-preserving models in which there are semantically significant approximate isomorphisms between structural properties of the representation and those of the represented event.

We are now in a position to see that experience replay in DQN involves representing past events in a way that is meaningfully similar to biological episodic memory. Like biological episodic memories, the representations used by DQN's 'experience replay' mechanism carry multidimensional information about events and do so in a manner that is at least partly iconic.

First, multidimensionality. Each state representation is an ordered sequence of observations and actions leading up to the time-step at which it is recorded. The observations represent the two-dimensional spatial properties of events. When stacked in ordered sequences, they additionally represent an event's internal temporal properties: the way that its spatial properties changed over time. By interspersing action representations between the observations in the sequence, state representations also represent the agent's involvement in the event, and how the environment changed in response to the agent's actions. *Experience* representations (the 4-tuples described in §2) combine state representations and action representations with reward representations, meaning that in total they carry information about the event's spatial and temporal properties, the agent's actions and how they affected the unfolding of the event, and how rewarding the event was for the agent. This seems like a reasonable approximation to the multidimensional information episodic memories prototypically carry: information about a remembered event's spatial and

temporal properties, about the subject's involvement in the event, and about their thoughts, feelings, or perceptions of the event.

Second, iconicity. State representations combine observations with action representations in an ordered sequence. As such, they are iconic in two ways. First, observations are iconic representations, in the sense that they are structured, their structural properties map onto the structural properties of their representata, and this mapping is semantically significant. Specifically, as noted in §2, each observation is formatted as a two-dimensional array of numbers, where the two dimensions represent the represented Atari screenshot's height and width, and each position in the numerical array corresponds to the pixel in the corresponding location in the screenshot. Second, combining these observations in sequence, together with action representations, produces another structured representation. In this representation, there is a correspondence between sequence position and time, such that items appearing earlier in the sequence are represented as having occurred at earlier points in time. The semantic significance of the structure of these representations means that they are unlikely to be retrievable piecemeal. Without the semantic information provided by the structured representation as a whole, a fragment of the representation would most likely be uninterpretable.

This is not to say that state representations are *wholly* iconic: in particular, luminance values and actions are represented in symbolic form. But this does not vitiate the claim that these representations are iconic in the ways I've outlined, since representations may combine both iconic and symbolic elements (Lee et al. 2022). We might construe a pixel vector as a hybrid map-like format: the numerical symbols stand for luminance values, whilst their location in the array stands for the locations of those values in the image. The use of symbols for luminance values does not negate the fact that there is a semantically significant structural correspondence between the array and the image. More importantly for our purposes, it is unlikely that episodic memories are wholly iconic: they are likely to involve some conceptual or semantic elements. Since our interest in this section is in the similarity between the representations involved in DQN's experience replay and those involved in episodic memory, a partially iconic representation of events with some symbolic components seems to fit the bill.

Episodic Memory's Function

In this section, I expand on the idea that by using DQN as a model of episodic memory, we might advance our understanding of episodic memory's function.

I noted in §1 that the function of episodic memory presents a puzzle for philosophers and scientists of memory. We might express the puzzle in terms of a twofold redundancy. First, given that we have other memory faculties including semantic memory, which stores general purpose information about the world, it seems redundant to have a faculty recording information about specific events. Second, much of

the information *stored* in episodic memory appears redundant, as it relates to highly specific events which will never be repeated and is unlikely to be directly useful in any other context. This redundancy, together with evidence that episodic memory is subject to systematic patterns of error and neurally overlaps with our faculty for imagining future and hypothetical scenarios, has led *simulationism* to become the dominant view of episodic memory's function (De Brigard 2014; Schwartz 2020). Simulationism is the view that episodic memory's function is primarily to support the imaginative construction of future and hypothetical scenarios.

Against this, several philosophers have recently mounted defences of the mnemonic view of episodic memory's function, on which its role is to store, encode, and retrieve information. For example, I've argued that episodic memory facilitates retrospective learning, that is, extracting novel information from an event after the event has passed (Boyle 2019). Simon Brown (2024) offers a related account, arguing that episodic memory supports *unrestricted* learning. The idea is that by capturing multidimensional information about events, episodic memory enables us to continuously revise and expand our models of the world. Elsewhere, I've also argued (Boyle 2021) that episodic memory plays a significant role in the storage, encoding and retrieval of semantic memory: it is both critically involved in the ordinary process of laying down semantic memories and provides internally generated cues for their retrieval. In a similar vein, Sara Aronowitz (2022) argues that we cannot understand the function of episodic memory in isolation. The process of semanticization, in which information from episodic memory becomes gradually more abstract and is encoded in semantic memory, suggests that episodic memory can be understood only in the context of a broader memory system.

Evaluating these theories is challenging. A natural way to approach the question would be to compare the behavioural repertoires of agents which have episodic memory with those of agents which lack it. We might do this by comparing humans with and without episodic memory deficits. This can be informative but faces some limitations: it can be difficult to know which behavioural differences are attributable to episodic memory. This is both because brain damage is rarely limited to *only* the brain areas involved in episodic memory, and because the brain areas supporting episodic memory may also support other cognitive functions: brain areas can multitask. Alternatively, we might try to compare the behaviours of animals with and without episodic memory. But the distribution of episodic memory in the animal kingdom is another significant puzzle: there is insufficient agreement about this for us to be sure which animals have or lack it.[6]

Given the resemblance between DQN and episodic memory at the algorithmic level, I propose that DQN and similar algorithms provide a testing ground for competing theories of episodic memory's function. In artificial agents, it is possible to ablate experience replay and to know that there has been no other intervention on

[6] For a discussion of this issue, see Boyle (2022).

the agent's cognitive architecture. Any differences in the agent's behavioural repertoire that result from this intervention can be traced directly to the agent's having or lacking the capacity to replay past events. It is also possible to tweak the inner workings of the experience replay algorithm and observe the effects of these changes. Insofar as artificial experience replay resembles biological experience replay at the algorithmic level—i.e., both provide integrated, partially iconic, multidimensional representations of events—this kind of investigation would provide defeasible evidence about the cognitive role function of episodic memory and might differentiate between accounts which are otherwise difficult to empirically evaluate.

Of course, DQN has not been used to directly test rival theories of episodic memory's function, so we should exercise caution in interpreting the results obtained with DQN in this way. Such caution notwithstanding, those results do suggest some support for mnemonic views, particularly those that emphasize episodic memory's role in semantic learning. In DQN, memories of specific events are used to train a network that learns more abstract, relational knowledge structures about the relationships between states of the environment and action/Q-value pairs. By using event memories in this way, DQN was able to learn much faster than rivals which do not make use of event memories in this way. It also attained a new state of the art in the Arcade Learning Environment, a set of complex problems involving high-dimensional sensory input. Disabling the experience replay component of the algorithm significantly impaired its performance. If we view the relationship between experience replay and the Q-learning network as analogous to the relationship between episodic and semantic memory, this provides some support to the theoretical claim that episodic memory supports the rapid acquisition of semantic memory. At the very least, it vindicates the idea that episodic memory *could* carry out a distinctive mnemonic function, even in the presence of a more general, abstract memory system. And given the similarities between DQN and episodic memory, I suggest, this provides some defeasible evidence that episodic memory carries out a similar function in us.

One might worry that the preceding argument overstates the similarity between DQN's experience replay and our episodic memory capacity. One reason for thinking this is that DQN records details accurately and the contents of its stored 'experiences' are not subject to change. By contrast, biological episodic memories are constructive and friable. We do not 'record' events accurately in entirely faithful detail. Our episodic memories are reconstructed at the time of retrieval, often drawing on general knowledge from semantic memory as well as details from the original event. As such, they are subject to change whenever they are retrieved, leading to systematic patterns of error (see De Brigard 2014 for discussion). A second, converse, issue is that biological episodic memories may include many details DQN's memories do not. Most of us remember perceptual and sensory information that goes beyond the visual: we might remember sounds, smells, sensations, and so on. Moreover, there are types of non-sensory content we recall as well, such as information about the emotions we felt or what we were thinking at the time of the event.

In summary: DQN's representations *include* a type and level of detail not typical of episodic memories, whilst also *excluding* some kinds of information characteristic of episodic memories.

However, using DQN to learn about episodic memory's function does not require that the two be exactly similar. Our interest is in using DQN as a *model* of episodic memory, such that we can draw justified inductive inferences about episodic memory by observing and manipulating DQN. Models can be useful in this way even if they are *simplified* or *idealized,* that is, even if they omit certain features of the modelling target, or include some features not found in the modelling target. As Catherine Stinson (2020) argues, what matters is whether both the model and the target belong to a common *kind* which licenses inferences from one to the other. In brief, what I have been arguing in this section is that DQN and episodic memory are members of a common kind: a kind of memory system characterized by the storage, encoding and retrieval of partially iconic representations storing detailed multidimensional information about past events. Moreover, our focal question about the function of episodic memory can reasonably be construed as a question about *this kind* of memory system: what is the use of a memory system which processes detailed, multidimensional representations of past events? So, despite the ways in which its representations differ from those involved in episodic memory, DQN seems like a promising model. Of course, it would take empirical work to establish the utility of this model; what I have been arguing is that there are good theoretical grounds for thinking that this empirical work would be fruitful.

Of course, there may be contexts in which these representational differences between DQN and episodic memory really matter. We might be interested in asking a somewhat narrower question about episodic memory's function, such as whether the reconstructive, error-prone processes that characterize our episodic memory confer epistemic or other advantages (Michaelian 2013; Puddifoot & Bortolotti 2018). In this context, a useful model would need to belong to a narrower kind: a *constructive* memory system processing similar representations of past events. DQN does not fall into this category. But this does not show that it would not be useful here, since the DQN algorithm could be adapted to incorporate constructive processes.[7] So, rather than vitiating the use of DQN as a model of episodic memory, the concern suggests how refinements to the algorithm might expand the range of theoretical questions with respect to which it is a fruitful model of episodic memory.

A related concern is that, notwithstanding the representational similarity between episodic memory and DQN, there remain significant differences at the algorithmic level relating to how these representations are processed. For example, in DQN, representations of past episodes are randomly sampled and used to train the Q-learning network. As I've indicated, semanticization in humans is a candidate analogue for this process: episodic memories are gradually consolidated and abstracted into

[7] For an example of a (non-DQN) episodic memory algorithm incorporating constructive processes, see Zakharov et al. (2020)

semantic memory. However, it is unlikely that our episodic memories are sampled *randomly* for this purpose: salience, recency and other factors are likely to have a significant impact on which memories are prioritized for semanticization. A related point can be made about forgetting. In DQN, the oldest stored experiences are erased when the episodic buffer reaches capacity. This does not mirror patterns of human forgetting: we forget many things besides our oldest memories and retain some memories for a very long time.

Again, there may be contexts in which these algorithmic level differences may not matter. If we're interested in understanding how a store of detailed, multidimensional event-specific memories can be used to support the acquisition of more abstract, general knowledge, DQN seems a suitable idealized model of this process. And as before, whilst there are contexts in which these differences do matter, the algorithm might be extended to resemble episodic memory more closely in relevant respects, so as to expand our understanding of episodic memory.

For example, Schaul et al. (2016) develop a DQN-based algorithm by adding prioritized experience replay. In this version of the algorithm, memories are not sampled randomly from the episodic buffer. Instead, memories with the highest temporal difference (TD) error are prioritized—that is, experiences which are more 'surprising' because the reward obtained differs significantly from the reward the system would predict. Alternative prioritization criteria could be used; the choice of TD error here is partly motivated by evidence that experiences with TD error are prioritized for replay in the hippocampus. This variant on DQN exhibited faster learning and a new state of the art in the Atari environment.

We might take this to provide defeasible evidence about the function of forgetting: at first blush, Schaul et al.'s results appear to support the view that forgetting is not a design flaw, but a critical design feature on a well-functioning memory system (Fawcett & Hulbert 2020; Michaelian 2011). Forgetting in biological systems can take two forms: information either becomes inaccessible or unavailable. Inaccessibility is a matter of information still being encoded in memory but being more difficult to retrieve; unavailability is a matter of the information having been entirely lost. We can see that deprioritization for replay in Schaul et al.'s algorithm provides an approximate analogue for inaccessibility: when information is deprioritized, it's less likely to be retrieved. As such, the algorithm suggests an adaptive role for at least one kind of forgetting: making events with low TD error less accessible leads to quicker learning and improved policies. Investigating the effects of different prioritization criteria might provide further insights into the function of both memory and forgetting.

As Schaul et al. note, one way to extend their work would be to apply prioritization criteria to erasure, for example, by erasing the memories with the lowest TD error, rather than the oldest memories, when it reaches capacity. Extending the algorithm in this way might shed light on the other form of forgetting: unavailability. In this vein, Ruishan Liu and James Zou (2019) investigate the relationship between the size of the memory buffer and learning rates, finding

that learning rates slow when the buffer is either too large or too small. They develop an algorithm in which the size of the memory buffer adaptive changes. If the TD error of the oldest memories is increasing, suggesting that these memories are becoming more informative, the buffer size increases and these memories are retained for longer. On the other hand, if the TD error of the oldest memories is decreasing, suggesting that they are becoming less informative, the buffer size decreases and these less informative older memories are more likely to be erased. Again, this suggests an adaptive role for prioritized patterns of forgetting in learning.

Similarly, we might note that whilst DQN only uses recorded experiences to train its Q network, our episodic memories are clearly put to other uses: most obviously, we frequently retrieve salient episodic memories to inform online decision making. This marks another significant difference in how episodic memories are processed in DQN and episodic memory. Once again, this simply shows that DQN is not a useful model in all contexts, as well as highlighting a way in which we might wish to develop experience replay algorithms to suit particular theoretical goals in cognitive science. If our interest is in investigating episodic memory's role in decision making, we would be better off looking at an algorithm in which recorded experiences are used in a similar way. For example, Blundell et al. (2016) develop an alternative experience replay algorithm they call the 'Episodic Controller'. In this architecture, past experiences are stored in a buffer which the agent can query to inform its decision making. When faced with a decision, the agent uses this body of stored knowledge to determine which actions have previously been associated with the highest reward in situations similar to the one it currently faces. The Episodic Controller learns quickly, especially in the early stages of confronting a novel problem, and particularly in sparse reward environments. In these scenarios, it exhibits behaviour 'akin to one-shot learning' (Blundell et al. 2016: 7). This architecture could be a fruitful model of episodic memory for the purposes of developing and evaluating accounts of episodic memory's role in decision making. At first blush, the view suggested seems to be that having access to information about individual, salient episodes facilitates fast learning in novel environments when reward is scarce.

Conclusion

Episodic memory presents many puzzles for memory theorists. Among them is the question of its cognitive role function: what does episodic memory contribute to the cognitive systems of which it's a part? I've argued that this question properly cast at the algorithmic level: what is the purpose of a memory system that exploits detailed, multidimensional, partially iconic representations of specific past events? It is difficult to gain empirical traction on this question by looking at biological systems. But, I've argued, DQN provides empirical leverage on the question in virtue of its similarity to episodic memory at the algorithmic level. In particular, against

increasingly popular simulationist views, the results obtained with DQN suggest that episodic memory plays a distinctive *mnemonic* role in the process of semantic learning, as several theorists have recently argued. Of course, there are significant differences between DQN and episodic memory in biological systems. But these do not undermine the utility of DQN in this context. Rather, they suggest ways in which we might adapt DQN or similar algorithms in future work, in order to evaluate theories about episodic memory's operations and its distinctive contributions to cognition.

References

Andonovski, N. (2022). Episodic representation: A mental models account. *Frontiers in Psychology*, 13, 1-23. https://doi.org/10.3389/fpsyg.2022.899371

Aronowitz, S. (2022). Semanticization challenges the episodic–semantic distinction. *The British Journal for the Philosophy of Science*. https://doi.org/10.1086/721760

Bellemare, M. G., Naddaf, Y., Veness, J., & Bowling, M. (2013). The Arcade Learning environment: An evaluation platform for general agents. *Journal of Artificial Intelligence Research*, 47, 253–79. https://doi.org/10.1613/jair.3912

Blundell, C., Uria, B., Pritzel, A., Li, Y., Ruderman, A., Leibo, J. Z., Rae, J., Wierstra, D., & Hassabis, D. (2016). *Model-Free Episodic Control*.

Boyle, A. (2019). Learning from the past: Epistemic generativity and the function of episodic memory. *Journal of Consciousness Studies*, 26(5–6), 242–51. https://doi.org/10.17863/CAM.35867

Boyle, A. (2020a). Remembering events and representing time. *Synthese*. https://doi.org/10.1007/s11229-020-02896-6

Boyle, A. (2020b). The impure phenomenology of episodic memory. *Mind & Language*, 35(5), 641–60. https://doi.org/10.1111/mila.12261

Boyle, A. (2021). The mnemonic functions of episodic memory. *Philosophical Psychology*, 1–23. https://doi.org/10.1080/09515089.2021.1980520

Boyle, A. (2022). Episodic memory in animals: optimism, kind scepticism and pluralism. In A. Sant'Anna, C. J. McCarroll, & K. Michaelian (eds.), *Current Controversies in Philosophy of Memory* (pp. 189–205). Abingdon: Routledge.

Brown, S. A. B. (2023). Episodic memory and unrestricted learning. *Philosophy of Science*, 91(1), 90–110. https://doi.org/10.1017/psa.2023.16

Clayton, N. S., Bussey, T. J., & Dickinson, A. (2003). Can animals recall the past and plan for the future? *Nature Reviews Neuroscience*, 4(8), 685–91. https://doi.org/10.1038/nrn1180

Clayton, N. S., & Russell, J. (2009). Looking for episodic memory in animals and young children: Prospects for a new minimalism. *Neuropsychologia*, 47(11), 2330–40. https://doi.org/10.1016/j.neuropsychologia.2008.10.011

Clayton, N. S., Yu, K. S., & Dickinson, A. (2001). Scrub jays (Aphelocoma coerulescens) form integrated memories of the multiple features of caching episodes. *Journal of Experimental Psychology: Animal Behavior Processes*, 27(1), 17–29. https://doi.org/10.1037/0097-7403.27.1.17

Crystal, J. D., Alford, W. T., Zhou, W., & Hohmann, A. G. (2013). Source memory in the rat. *Current Biology*, 23(5), 387–91. https://doi.org/10.1016/j.cub.2013.01.023

Crystal, J. D., & Smith, A. E. (2014). Binding of episodic memories in the rat. *Current Biology*, 24(24), 2957–61. https://doi.org/10.1016/j.cub.2014.10.074

Cummins, R. (1975). Functional analysis. *The Journal of Philosophy*, 72(20), 741–65.

De Brigard, F. (2014). Is memory for remembering? Recollection as a form of episodic hypothetical thinking. *Synthese*, 191(2), 155–85. https://doi.org/10.1007/s11229-013-0247-7

Eacott, M. J., Easton, A., & Zinkivskay, A. (2005). Recollection in an episodic-like memory task in the rat. *Learning and Memory*, 12(3), 221–23. https://doi.org/10.1101/lm.92505

Fawcett, J. M., & Hulbert, J. C. (2020). The many faces of forgetting: Toward a constructive view of forgetting in everyday life. *Journal of Applied Research in Memory and Cognition*, 9(1), 1–18. https://doi.org/10.1016/J.JARMAC.2019.11.002

Gershman, S. J., & Daw, N. D. (2017). Reinforcement learning and episodic memory in humans and animals: An integrative framework. *Annual Review of Psychology*, 68, 101–28. https://doi.org/10.1146/annurev-psych-122414-033625

Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2017). Neuroscience-inspired artificial intelligence. *Neuron*, 95(2), 245–58. https://doi.org/10.1016/j.neuron.2017.06.011

Lee, A. Y., Myers, J., & Rabin, G. O. (2022). The structure of analog representation. *Nous*. https://doi.org/10.1111/nous.12404

Levine, B., Svoboda, E., Hay, J. F., Winocur, G., & Moscovitch, M. (2002). Aging and autobiographical memory: Dissociating episodic from semantic retrieval. *Psychology and Aging*, 17(4), 677–89. https://doi.org/10.1037//0882-7974.17.4.677

Liu, R., & Zou, J. (2019). The Effects of Memory Replay in Reinforcement Learning. *2018 56th Annual Allerton Conference on Communication, Control, and Computing, Allerton 2018*, 478–85. https://doi.org/10.1109/ALLERTON.2018.8636075

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco, CA: W. H. Freeman.

Michaelian, K. (2011). The Epistemology of Forgetting. *Erkenntnis*, 74(3), 399–424. https://doi.org/10.1007/s10670-010-9232-4

Michaelian, K. (2013). The information effect: Constructive memory, testimony, and epistemic luck. *Synthese*, 190, 1–28. https://doi.org/10.1007/s11229-011-9992-7

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G. et al. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529–33. https://doi.org/10.1038/nature14236

Puddifoot, K., & Bortolotti, L. (2018). Epistemic innocence and the production of false memory beliefs. *Philosophical Studies*. https://doi.org/10.1007/s11098-018-1038-2

Rubin, D. C., & Umanath, S. (2015). Event memory: A theory of memory for laboratory, autobiographical, and fictional events. *Psychological Review*, 122(1), 1–23. https://doi.org/10.1037/a0037907

Schaul, T., Quan, J., Antonoglou, I., Silver, D., & Deepmind, G. (2016). *Prioritized Experience Replay* (pp. 1–21).

Schwartz, A. (2020). Simulationism and the Function(s) of Episodic Memory. *Review of Philosophy and Psychology*, 11(2), 487–505. https://doi.org/10.1007/s13164-020-00461-1

Shea, N. (2014). Exploitable isomorphism and structural representation. *Proceedings of the Aristotelian Society*, 114(2), 123–44. https://doi.org/10.1111/j.1467-9264.2014.00367.x

Stinson, C. (2020). From implausible artificial neurons to idealized cognitive models: Rebooting philosophy of artificial intelligence. *Philosophy of Science*, 87(4), 590–611. https://doi.org/10.1086/709730

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction*, 2nd Edition. Cambridge, MA: MIT Press.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of Memory* (pp. 381–402). New York, NY: Academic Press.

Tulving, E. (2005). Episodic memory and autonoesis: Uniquely human? In H. S. Terrace & J. Metcalfe (eds.), *The Missing Link in Cognition* (pp. 4–56). New York, NY: Oxford University Press.

Zakharov, A., Crosby, M., & Fountas, S. (2020). Episodic memory for learning subjective-timescale models. https://arxiv.org/abs/2010.01430v1

12

Memory and Planning in Brains and Machines

Multiscale Predictive Representations

Ida Momennejad

Imagine planning an overseas trip. You could consider the plan in exhaustive detail, down to which seat you prefer, what you will pack, your commute plan, which socks to wear. Or you could consider the most crucial steps, thinking about it at a high level by jumping over many steps and merely considering key steps such as getting a plane ticket and booking a hotel. In this imaginary scenario you easily use representations of the world that are organized in your memory at multiple scales. This example illustrates the central thesis of this chapter. The ability to move back and forth across representations of the world, to remember the past and imagine and plan the future at different scales of abstraction, is crucial to the structure of memory and its use in prediction and planning.

What should these memory structures look like to enable planning and how do we study them? This chapter reviews evidence showing that the learnt structures of events, such as the spatial structure of an environment (Figure 12.1) or the relational structure of a social network, are organized in memory as predictive representations that are multi-step and multiscale. That is, brain regions can organize the same memories with different predictive horizons or scales (Figure 12.1). The scale varies depending on the task at hand or the level of abstraction involved. Much of this evidence is derived from experimental areas where the roles of memory in planning are particularly evident, including spatial navigation and non-spatial associative inference.

What is a cognitive map? A cognitive map is a representation of latent relational structures that underlies a task or environment. It is how our knowledge or memories are organized in order to facilitate efficient retrieval, planning, reasoning, and inference in biological and artificial agents (Tolman 1948; Kumaran and Maguire 2005; Epstein et al. 2017; Behrens et al. 2018; Momennejad 2020; Brunec and Momennejad 2021).

The concept originated from Tolman's latent learning experiments, demonstrating rodents' ability to learn the latent structure of a maze without rewards (Tolman 1948). This challenged the dominant behaviourist dogma of the time that learning

Ida Momennejad, *Memory and Planning in Brains and Machines*. In: *Space, Time, and Memory*. Edited by: Lynn Nadel and Sara Aronowitz, Oxford University Press. © Oxford University Press (2025). DOI: 10.1093/oso/9780192882547.003.0012

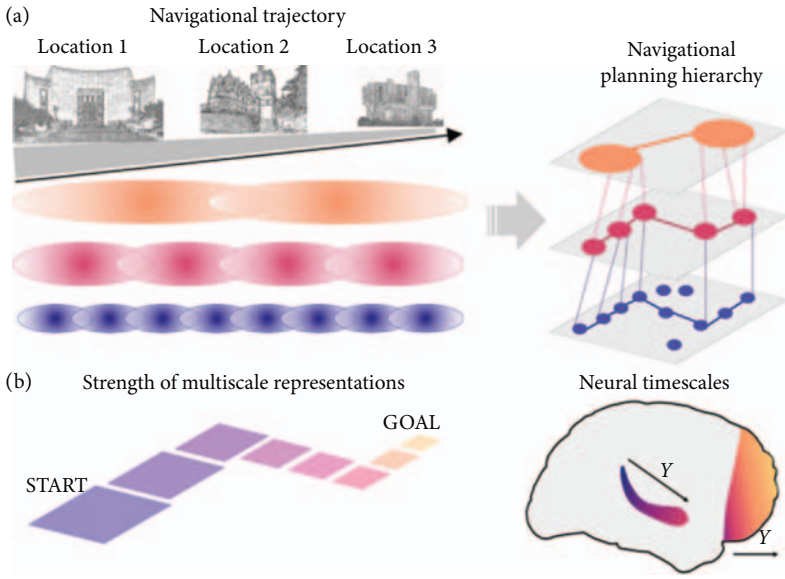


Figure 12.1 Multiscale predictive representations in navigation and planning. (A) A navigational trajectory can be represented with varying granularity: e.g., step by step (purple), block by block (pink), or subgoal to subgoal (orange). (B) During planning, the scope or horizon 'visible' that each scale covers can vary from a step to the goal. Here, review evidence that these gradients of scales are reflected in representational hierarchies in the hippocampus and the prefrontal cortex, and are called on flexibly in navigational memory and planning.

only occurs with reinforcement; and paved the way for a cognitivist revolution. Decades later, discoveries of hippocampal place cells (O'Keefe and Dostrovsky 1971; O'Keefe 1976; O'Keefe and Nadel 1978) and entorhinal cortex grid cells (Fyhn et al. 2004; Hafting et al. 2005; Moser, Kropff, and Moser 2008), together referred to as 'the brain's GPS', further substantiated cognitive maps and earned the 2014 Nobel Prize (The Nobel Prize in Physiology or Medicine 2014 2014).

Cognitive maps have since been studied behaviourally, computationally, and neurally. More recent studies suggest that multi-step, multiscale, and compressed neural representations are crucial for inference in both memory and planning (Behrens et al. 2018). Over the past decades, a number of Reinforcement Learning (RL) and deep neural network models have been proposed to capture the computations involved in cognitive maps and planning, especially in the hippocampus and the prefrontal cortex of humans, rodents, bats, monkeys, and birds (Epstein et al. 2017; Behrens et al. 2018; Brunec and Momennejad 2021; Pudhiyidath et al. 2022).

How does your brain organize cognitive maps? *How* is the organization of memory (the past) connected to prediction and planning (the future)? One possibility is that the brain may store and unroll a step-by-step map of the environment for planning. This is similar to the idea of model-based reinforcement learning (MBRL) (Sutton and Barto 2018), which we will discuss in the following section.

Note that MBRL is indeed a method that allows for multi-step planning, and the options framework allows for multi-step compression (Sutton et al. 2023; Sutton, Precup, and Singh 1999). However, the point of discussion here is whether the *representation* of the environment, or the map of its states, is stored as multi-step compressed representations, or as one-step tuples that can be unfolded for planning.

Another possibility is that the representational structure of memory is both multi-step and multiscale. This allows for 'jumps' over multiple steps in memory, i.e., how many steps a compressed representation 'jumps' over. Here the horizon or 'scale' of these jumps rely on compression and the predictive nature of the representation at a given scale. Similarly, computational and empirical studies of cognitive maps suggest that they are organized as predictive representations that are also multiscale, organized with different predictive scales, or horizons (Figure 12.1). While there are various approaches to temporal abstraction, one idea highlighted in this chapter involves the successor representation (SR) in reinforcement learning, proposed by Dayan, and variations that combine SR with replay and propose multiscale successor representations (Dayan 1993; Momennejad et al. 2017; Momennejad 2020; Machado et al. 2023). We will discuss this in more detail in sections 1 and 2.

Accumulating evidence over the past decades is consistent with the idea that such predictive representations may govern human behaviour in episodic memory tasks (Gershman et al. 2012), planning and decision-making (Momennejad et al. 2017; Russek et al. 2017), and further, that these compressed representations may support cognitive maps (Behrens et al. 2018) in the rodent hippocampus and entorhinal cortex (Stachenfeld, Botvinick, and Gershman 2017; Geerts et al. 2020; de Cothi et al. 2022).

Consistently, a recent human fMRI study showed that naturalistic VR navigation relies on multiscale predictive representations in prefrontal and hippocampal hierarchies (Brunec and Momennejad 2021). This study showed that weighted representational similarity at longer scales are associated with goal-directed naturalistic navigation. Consistently, another study suggests the brain uses hierarchical representations in anticipation of events that are multiple steps away (Tarder-Stoll, Baldassano, and Aly 2023). A study of complementary task representations in hippocampus and prefrontal cortex showed that task abstractions in medial prefrontal cortex simultaneously represent behaviour over different temporal scales (Samborska et al. 2022). More evidence discussed in section 3 onward.

In this chapter I will first establish how temporal abstraction connects the past and the future, then introduce the main computational methodology (reinforcement learning, RL) for framing this connection in terms of predictive representations at multiple scales, and then discuss a number of behavioural, neural, and computational studies that support the computational idea. I will end by discussing why these approaches can inspire novel ways to evaluate and augment planning and navigation abilities of generative AI.

Temporal Abstraction: Binding the Past and Future

Consider how we come to learn representations of the world. One possibility is that as we navigate the world our brains store each event that we experience individually, and later on retrieve them individually. In this framework, planning requires rolling out and evaluating every single event step by step to identify the optimal planning trajectory. This is similar to model-based RL (MBRL, Figure 12.2).

Let us consider another hypothesis about how brains learn, store, and update representations of the world. As we navigate environments, each event leaves various traces in our memory systems, and these traces last for variable durations, their traces decaying at different rates and scales in different parts of the brain. Over time the trailing traces of events that occurred closer in time overlap, leading to an association among events that are closely associated.

Gradually these overlapping traces are consolidated as associations that, depending on the horizon and decay rate of the trace, lead to temporal clustering of events that occurred within overlapping horizons (Figure 12.1). This is also a form of compression, or temporal abstraction, which can further lead to the efficient storage of memories, categorization, or the ability to plan. The primary scales or horizons may depend on the size and complexity of the environment in spatial settings, or on the frequency of surprising events more broadly for both spatial and non-spatial associative settings.

Now let us consider how the overlap of trailing traces and the ensuing temporal abstractions serve prediction. Let us say the memory of event A (Figure 12.1A, location 1) occurred at time t1, and event B (Figure 12.1A, location 2) at time t3, and event C (Figure 12.1A, location 3) at time t5. Now consider a part of the brain where the trace of each event lasts for three time steps. As such, each time the memory of event A is triggered by the world, we get a partial activation of event B even when it is not present, because the memory of event A is already bound to the activation of event B so each time A is activated, B is partially activated too.

As coactivation of successor events occurs over time, an event's activation leads to the partial activation of its successor state in a predictive manner (Figure 12.1, B). The strength and scope of the chain of successor event activation depends on two factors. First, it depends on the *distance* of each state from the current one. Second, it depends on the horizon of temporal compression in the part of the brain where these representations are activated.

Evidence from human fMRI (Figure 12.2) suggests that the scale or horizon of the temporal abstraction increases along the posterior to anterior axis of the hippocampus and entorhinal cortex (Strange et al. 2014; Nielson et al. 2015; Collin, Milivojevic, and Doeller 2015), as well as the rostro-caudal axis of the prefrontal cortex during *goal-directed navigation* (Brunec and Momennejad 2021; Patai and Spiers 2021), *multi-step planning* in sequential and hierarchical problem solving (Botvinick and Weinstein 2014), and *prospective memory* (Burgess et al. 2008; Momennejad and Haynes 2012; Momennejad and Haynes 2013).

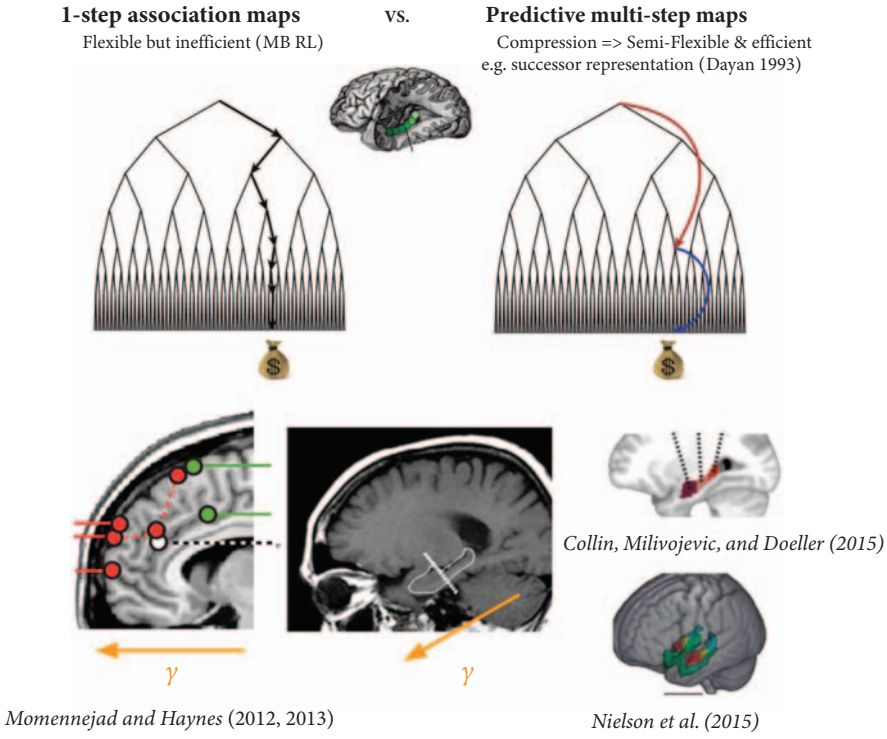


Figure 12.2 What do cognitive maps look like? (Top) Two computational hypotheses about the structure of cognitive maps in memory. One possibility is that the brain stores step-by-step (1-step) associations and unrolls them for evaluation at the moment of decision-making. This is similar to the transition probabilities in model-based RL (Sutton and Barto 2018). Another possibility is that the brain may store jumpy multi-step associations using *temporal abstraction*, simplifying prediction, and planning. Computationally, this is the successor representation (Dayan 1993). (Bottom) Evidence from human fMRI suggests that predictive representations may be learned at different scales along the hippocampal and prefrontal hierarchies such that more anterior regions are more likely to represent larger scales, γ, corresponding to more temporal or spatial abstraction.

The above-mentioned hypotheses can be formalized in the service of computational modelling and quantitative analysis of representations underlying memory, navigation, and planning. Section 2 considers a common approach for such a formalization, the computational framework of reinforcement learning.

The Computational Approach: Framing Memory and Predictive Representations with Reinforcement Learning

Reinforcement learning (RL) is a computational framework that simplifies value-based and goal-directed decision-making and planning behaviour. The RL framework posits an agent in an environment with a number of states, such that from each non-terminal state S the agent can take action, a, to move to the next state S' and receive reward r (Sutton and Barto 2018).

In classic RL, the agent's goal is often simplified as maximizing expected value in this environment. Most RL algorithms aim to capture how an agent achieves the goal of maximizing value in exploration, learning, and decision-making phases. That is, first, how RL captures how the agent takes actions in the environment to explore states, learn the state values or state-state relationships and rewards. Second, RL captures how the agent combines this knowledge to find the policy or sequence of actions, $\pi^{*} or\, \pi(a|S)$ that maximizes value. This is known as optimal policy, leading to optimal value or V^{*}.

This simple computational framework offers various algorithms for learning representations with abstraction, hierarchy, generalization, and transfer. The most commonly known RL algorithms within the cognitive and neurosciences include model-free RL, model-based RL, the successor representation, and hybrid algorithms combining one or more elements of these algorithms with each other and with offline replay. Let us take a closer look at them.

Model-Free RL, Model-Based RL, and the Successor Representation

Briefly, model-free agents simply learn to cache the value of each state and action, *Q(S, a)*, in a lookup table or a value function *V(S)*. Importantly this approach combines states, actions, and rewards to cache a scalar expected value without storing state-state relationship, i.e., the probabilities of states leading to one another. Since model-free RL does not store representations of the relational structure of states, it *can not* serve as a candidate for formalizing the structure of cognitive maps and memory. The approach has been fruitful in the study of simpler value-based decision-making and 'habits' (Gläscher et al. 2010), and in modelling striatal learning in the dual-systems hypothesis (Daw, Niv, and Dayan 2005).

Let us consider an RL algorithm that can serve as models of memory. Model based RL agents (MBRL) capture the relationship between adjacent states, that is, the probability of transition to a state S' given the agent starts in state S and takes action a. This is formalized as *p(S' | S, a)*. As such, MBRL stores the reward for state and action tuples, r*(S, a)*, on the one hand; and the relational structure of the environment, *p(S' | S, a)*, on the other, i.e., a map of one-step transition probabilities (Figure 12.2, Table 12.1).

Unlike model-free RL, the model-based agent does not use a look-up table of cached values to determine the optimal policy (or the sequence of actions that yield the highest value given the goals). Instead, MBRL rolls out, step-by-step, all possible policies iteratively, computing the expected discounted value of each trajectory and picking the policy, or sequence of actions, that maximizes rewards. Updating the value function relies on iterations using the Bellman equation:

$$Equation 1. V^{*}(S) = max_{a}[R(S,a) + \sum_{S'} p(S'|S,a)\gamma V^{*}(S')]$$

The Bellman equation is fundamental to RL. Here we only discussed the Bellman equation in relation to MBRL. However, note that the Bellman equation can be used to estimate model-free value as well. For MF value estimation, the equation does not include the transition probabilities among states, p(S′ |S, a), but the value of the next state is either based on the actual action taken and direct experience as in Sarsa (on-policy method):

$$Q(s,a) = Q(s,a) + \alpha^*[R + \gamma^* Q(s',a') - Q(s,a)]$$

or estimated based on the maximum Q-value of the next state regardless of the action taken in Q-learning (off-policy algorithm):

$$Q(s,a) = Q(s,a) + \alpha^*[R + \gamma^* max(Q(s',a')) - Q(s,a)]$$

So far a candidate model of memory we discussed is model-based RL, with the state transition probability T as a potential model of the relationship among states in memory. In Figure 12.2 (top left) we considered such a one-step tuple organization (Daw et al. 2011) as one of the hypotheses about how cognitive maps or relational structures may be organized in memory.

We also considered another possibility (Figure 12.2, top right). Namely, that representations of events and their relationships, or cognitive maps, can be organized beyond one-step relations: as predictive representations of successor states in memory. The RL framework offers a formalization of this perspective in terms of the successor representation (Dayan 1993).

The successor representation (SR) caches multi-step relationships among states. Specifically, it captures how often on average an agent expects to visit a successor state *S*' starting in state *S* and moving under policy, π. Unlike model-based RL, the SR agent doesn't store probabilities of state–action–state transitions. The SR between a given state and a successor state (that can be multiple steps away) is computed by multiplying a discount parameter and the probability of state–action–state transition at every state (see Equation 2).

One way to construct the SR matrix, which we will call *M* following Dayan's formulation (Dayan 1993), is to derive it analytically from the transition probability of a model-based agent's *T*, where matrix *T* stores the state-by-state transition probabilities, or the expected transition distribution under the policy. The equation would be as follows:

$$\textit{Equation 2. } M = \sum_{t=0}^{\infty} \gamma^t T^t = (I - \gamma T)^{-1}$$

$$\textit{Where value for state S is computed as: } V(S) = \sum_{S'} M(S,S')\,R(S')$$

Intuitively, Equation 2 is derived by summing discounted probable states over time, $M = \sum_{t=0}^{\infty} \gamma^t T^t$, which mathematically converges to $(I - \gamma T)^{-1}$.

However, we do not need to learn T first in order to learn the SR matrix. The SR can be learned via temporal difference learning, in which a row of SR, *M(S)*, is updated as the agent moves from state *S*, as follows:

$$Equation\ 3.\ M(S) = M(S) + \alpha\left(onehot(S) + \gamma M\left(S'\right) - M(S)\right)$$

This formulation allows for the gradual learning of the successor representation over time (Dayan 1993).

M can be initiated as a matrix of zeros or the identity matrix, when each state could lead to itself. Then, each time the agent visits state *S*, the *Sth* row of M is updated according to Equation 3: with learning rate α, discount parameter γ, and a successor prediction error. At every observation of a state, action, state transition, the *successor prediction error* is the sum of a onehot vector, all zeros except for index S, and a discounted *M(S')* or the S'th row of the successor representation, minus the previous *M(S)*. Note that the vector of rewards R is stored separately.

At the moment of decision, unlike model-based RL, the SR agent does not need to roll out every single possible path and decide which one is best. Instead of iterating Equation 1, the expected value can be computed via a linear combination or dot product of *M* and *R* (the reward vector). In a sense, SR 'pre-bakes' the unrolled combination of probabilities and discount parameters over time, compared to the one-step structure stored by a tabular[1] model-based agent. Therefore, by merely multiplying the reward vector, the SR agent can come up with a decision faster than MBRL.

A growing number of studies find evidence for the successor representation, and its eigenvectors, in memory and decision-making tasks, alluding to the possibility that it may serve as a general principle for the organization of memories (Gershman et al. 2012; Schapiro et al. 2013; Stachenfeld, Botvinick, and Gershman 2017; Momennejad et al. 2017; Garvert, Dolan, and Behrens 2017; de Cothi and Barry 2020; Bellmund et al. 2020; Momennejad 2020).

Flexible Behaviour as a Window Into Latent Representations

Imagine you are on the way to work. You notice your favourite food truck is unusually parked a few blocks to the west of the office. During lunch will you default to habit and go three blocks south, where the food truck usually parks, or will you integrate that morning's observation with your knowledge of the city and go west?

An important aspect of planning is flexibility of behaviour in response to local changes in the environment that require integrating past memories with new

[1] Note that deep MBRL such as DreamerV2 (Hafner et al. 2020) can have representations that go beyond one-step. That said, it's been shown that deep MBRL behaviour does not pass tasks requiring the flexibility expected of tabular MBRL (Wan et al. 2022). That said, here we follow a neuroscience tradition that focuses on the tabular notion of MBRL (Daw, Niv, and Dayan 2005; Daw and Dayan 2014; Daw et al. 2011).

experience. There are at least two kinds of local changes, both of which require flexibility of planning. This could be a change in the location of rewards (like the food truck), or a change in the transition structure (e.g., a road is blocked, requiring a detour, or F train is running on the A train track, etc.).

Earlier we discussed RL agents that offer candidate models of memory, namely MBRL and varieties of SR-based algorithms.[2] To address the connection between memory and planning, a reasonable question follows. *Which agent's behaviour on planning tasks is more similar to humans?* One way to discern whether human behaviour is more similar to a model-based or SR approach is to design experiments that elicit different behaviour in these agents. One such example are experimental paradigms that probe flexibility to local changes in rewards and transition structures, e.g., blocked road, swapped train track (see section 2c for detail).

In the face of a local change (like the food truck example above) model-based RL only needs to have updated the one-step change. If this update was successful while experiencing the local change, MBRL can solve the problem with the same reaction time as planning without such a change. The SR agent, on the other hand, only learns from direct experience (according to Equation 3) and, therefore, can adapt to local changes in rewards (where the structures don't change) but, without using offline replay, is worse at adapting to local changes in transition structures. For instance, it would update the row of the SR associated with the truck but not the row associated with exiting the office. This can be mitigated if the agent could replay the trajectories related to the change it has observed.

This idea that SR could be updated *offline* according to Equation 3 has been introduced in terms of *SR-Dyna*—or *DynaSR* (Barnett and Momennejad 2022)—and hybrid SR–MB (Momennejad et al. 2017; Russek et al. 2017). The replay component, Dyna, was proposed in Rich Sutton's proposal of the Dyna architecture, where a model-free agent's policy is updated offline via simulated experience or roll-outs of the model-based T (Sutton 1991).

In sum, we have three hypotheses of how latent representations of an environment may be organized when humans engage in flexible planning behaviour. The cognitive map, or the relational structure of the environment may be stored as a *one-step map* (MB), a *multi-step map* (successor representation or SR) only learned and updated online, or a multi-step map that can be updated both *online and offline* (SR-Dyna or DynaSR).

The next section dives deeper into tasks that elicit different behaviour from these models: the retrospective revaluation of rewards (the structure doesn't change but the rewards change), the retrospective revaluation of transitions (the structure has changed but not the rewards), and empirical results comparing human and agent behaviour in these tasks.

[2] Recall that model-free agents cannot adapt to *local* changes without re-experiencing the entire trajectory after the change.

Table 12.1 Comparison of model representations, value computation, and behaviour. Both the MF cached value and the SR can be learned via simple temporal difference learning during the direct experience of trajectories in the environment. M, the SR or a 'rough' predictive map of each state's successor states; value function Q(s, a) (which maximizes value for optimal policy; compare to V(s), value function for a state under the current policy); R, reward function; T, full single-step transition matrix.

	Representation	Computation	Behavior
MF learner	*Q*: Cached value	Retrieve cached value Lowest cost	Habit, Fast
MB learner	*R*: Vector of all state rewards *T*: One-step state transitions matrix	Iteratively compute values Highest cost, resource-constrained	Fully flexible, Slow
SR learner	*R*: Vector of all state rewards *M*: Multi-step future state occupancy matrix (policy-dependent caching)	Combine cached future occupancies with rewards Intermediate costs	Semi-flexible, Fast
Hybrid SR	R & M (as above) SR output combined with T, or update SR with replay (on- or offline)	Combine SR with MB or replay Intermediate costs: mostly SR costs, at times MB or replay costs	Flexible but asymmetric, Fast (mostly)

Experimental Tasks to Compare Behaviour: Model vs. Human

Let us return to the central question: How are memory representations structured, and how do these structures connect to prediction and planning? One possibility we discussed is similar to a model-based RL agent that stores one-step transition probability tuples $p(S' \mid S, a)$ and then rolls out possible paths according to Equation 1 to determine the optimal one (Figure 12.2). Another possibility is that there are multi-step contingencies rather than one-step transitions. A common version of this is the successor representation or DynaSR, where SR is updated both online and offline.

As noted earlier these models respond differently to tasks with partial changes in rewards and transition structures. Such tasks require the integration of memory and new experience for planning. While model-free RL cannot solve any of the problems (Figure 12.3), model-based RL can solve both reward revaluation and transition revaluation symmetrically, online SR can only manage reward revaluation, and SR with replay or DynaSR can adapt to both changes. However, the latter needs more computation for transition revaluation, using offline replay to stitch together different pieces of the past to update the present policy.

In a series of experiments, researchers designed a number of simple flexible planning tasks (Figure 12.3) to measure and compare human and model behaviour.

Interestingly, human behavioural results reflect an asymmetry (Figure 12.3, bottom): while humans can solve both tasks, they significantly perform better at retrospective revaluation of rewards compared to transition revaluation. These findings were replicated a number of times in tasks with simple line graph structure and tree structure (Momennejad et al. 2017; Russek et al. 2021).

Taken together, human behavioural results are consistent with the DynaSR model, which predicts both asymmetric accuracies on reward and transition revaluation as well as longer reaction times for transition revaluation. Another behavioural result to note is that this asymmetry is not the result of a speed accuracy trade-off, that is worse performance is not accompanied by faster reaction times (RTs). In fact, on the contrary, human reaction times are significantly longer for solving transition revaluation (on which human performance was less accurate) compared to reward revaluation (Figure 12.3, bottom right).

Accounting for both differences in accuracies and reaction times is important for computational models of brain function. For instance, a recent study considered the hypothesis that a probabilistic successor representation may capture the same behavioural results. However, while the probabilistic approach captures the asymmetry in accuracies in its present format it cannot trivially account for reaction time differences observed in humans, without ad hoc additions and assumptions (Geerts, Stachenfeld, and Burgess 2019).

Methodologically, the importance of explaining various dimensions of behaviour (e.g., accuracy and RT) in computational models of cognitive capacities and brain function cannot be overstated. This is especially important in the field of neuroAI and building human-like agents (Momennejad 2023).

Prioritized Replay during Offline Learning

We discussed the importance of offline learning via replay as a key component of DynaSR, the model that best captured human behaviour in tasks that required integrating memory and planning (Figure 12.3). However, in real-life problem settings there are far too many memories to replay and reactivate, such that were brains to replay them at random, noticing the most relevant memories and stitching them together could become a matter of luck, or even intractable. So how can our models of brains account for choosing the most relevant memories to replay?

One possibility is that the model that best captures behavioural and neural findings *prioritizes* which memories to replay according to the amount of surprise associated with them. This notion of surprise can be captured in terms of prediction errors, computed as the difference between what the agent expects and what it observes, which can be signed (positive or negative prediction error) or unsigned (simply a surprising event). This idea has been used in reinforcement learning since

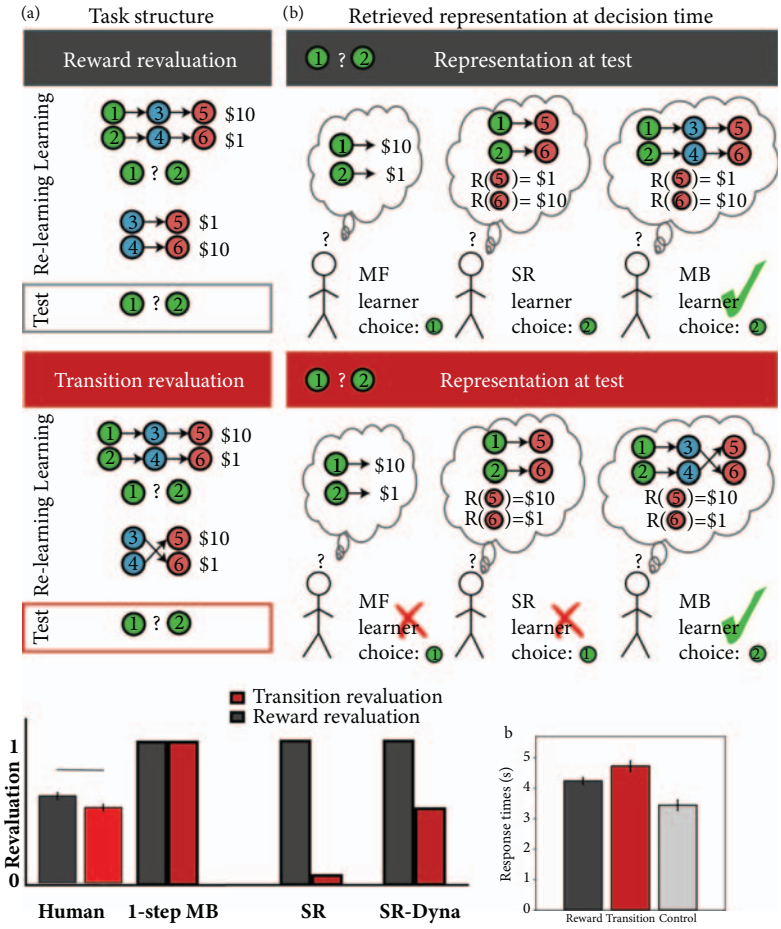


Figure 12.3 Experimental design, hypothesized representations, human, and model behaviour. (Top) The design of reward and transition revaluation and a schematic illustration of how each of the three RL agents discussed earlier would solve the problem. (Bottom) Model-free agents cannot solve either, model-based is expected to solve both equally well, a successor representation learner without replay is expected to solve reward revaluation but fail at transition revaluation, and a hybrid agent learning SR both online and offline via replay (SR-Dyna or DynaSR) is expected to learn both but be better at reward revaluation. Human behaviour in terms of both accuracy and reaction times is more consistent with DynaSR agents.

the 90s (Moore and Atkeson 1993) and still echoes in contemporary RL and deep learning algorithms as much as in psychology and neuroscience (Sutton et al. 2012; Schaul et al. 2015; Mattar and Daw 2018; Rouhani and Niv 2021).

While it is more common to consider reward prediction errors in the context of memory prioritization, it is also possible to consider successor prediction errors (PE) as discussed earlier in relation to Equation 3. A recent modelling paper proposed PARSR, Priority Adjusted Replay for Successor Representations, offering variants of DynaSR where the replay priority can be set to reflect *reward PE* or

successor PE (Barnett and Momennejad 2022). The study found differences in the representations learned by the different prioritization approaches, which may better serve different types of tasks. Future research is required in this area to better elucidate the usefulness of these different approaches.

Neural Representations: Model vs. Human

Evidence for Predictive Representations

We discussed behavioural evidence for the hypothesis that cognitive maps may be organized similarly to the successor representation: predictive representations of expected visitations to successor states. SR has a number of properties that lend well to a number of testable neural predictions. For instance, the rows of SR capture the future, or the successors of each state, and the columns of which capture the past, or the predecessors of each state (Figure 12.4).

Importantly, it has also been shown that the columns of SR resemble the representation of place fields in the hippocampus, and explain observed phenomena including backwards expansion on tracks and elongation near walls (Stachenfeld, Botvinick, and Gershman 2017; George et al. 2022). Note that policy-dependent (or path-dependent) SR representations predict that the brain represents distances not in Euclidean terms, but in a path-dependent fashion, see Figure 12.4 (Russek et al. 2017; Momennejad 2020).

This is in line with a number of observations in rodent and human neuroscience. Mehta and colleagues have shown that place fields become more predictive with experience, i.e., they expand along the direction of the path experienced (Mehta, Barnes, and McNaughton 1997; Mehta, Quirk, and Wilson 2000), see the rightmost representation in Figure 12.4. Moreover, distance to goals has been shown to be path dependent and not Euclidean in human medial temporal lobe (Spiers and Maguire 2007; L. R. Howard et al. 2014). This is in line with a multiscale SR model discussed later, the derivative of which can capture distance among states (Momennejad and Howard 2018), see section 3d for more detail.

A human fMRI study (Russek et al. 2021) conducted the flexible planning experiment discussed in section 2c (Momennejad et al. 2017) in the scanner. The study focused on transition revaluation (Figure 12.3), investigating whether fMRI evidence for outdated successor representations (i.e., persisting representational similarity of a starting state to a further state that is no longer a successor) could predict behavioural errors in transition revaluation. They found that indeed, in regions associated with navigation and planning, representational similarity to outdated successors predicted behavioural errors (Russek et al. 2021).

If the SR captures the structure of representations that underlie navigation, some further behavioural predictions could be made. A recent paper specifically investigated how deforming the cognitive map may distort memories and showed how a

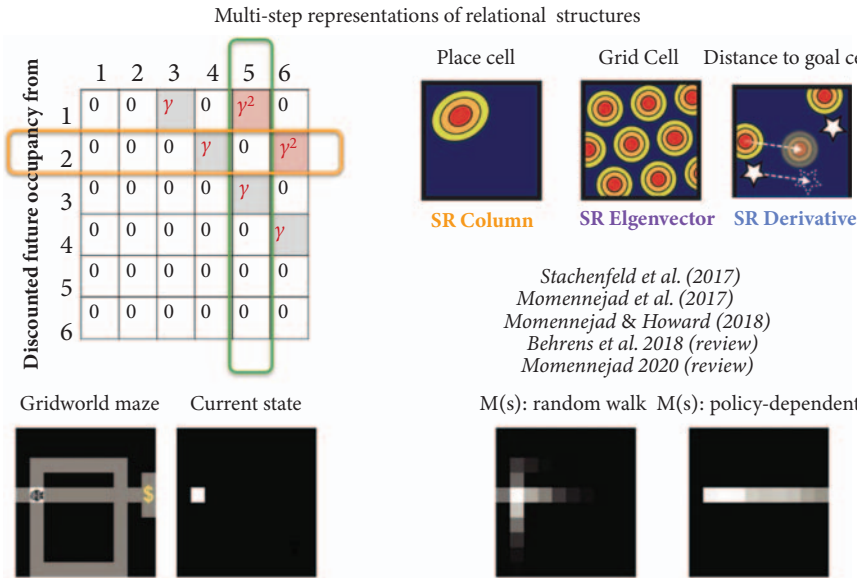


Figure 12.4 Successor representation as a model of predictive cognitive maps. (Top Left). A successor representation matrix for the graph structure of the task in Figure 12.3 is displayed. The rows correspond to the successor states that are one or more steps away, or the future (e.g., the successors of state 2 have non-zero successor representations in row 2). The columns represent the predecessors of a state (or the past). NB: the diagonal can also be 1, assuming there's a possibility of going to a state from that state. (Top Right) Evidence suggests that columns of SR simulate place fields in the hippocampus, while the eigenvector of SR simulates grid fields, and the derivative of multiscale SR estimates the distance among states. These findings make SR, a path or policy-dependent representation, a viable computational candidate for capturing memory and space in the medial temporal lobe (MTL). (Bottom) Consider a rodent in the gridworld maze. Its current location, its successor representation according to a random walk (independent of the path it takes), and according to the policy or path it takes are displayed. Place fields become predictive of the path with experience.

model with SR and its eigenvectors can explain the phenomenon (Bellmund et al. 2020). The study used an innovative design, in which the structure of the environment goes from a rectangle to a trapezoid, leading to distortions in behavioural judgements such as distance.

Another study proposed a model of how environmental manipulations may impact predictive representations in the medial temporal lobe. They proposed a model where the SR is learned from a basis set of boundary vector cells (BVCs), and showed that the model captures place cell firing in terms of successor features, while grid cells represent a low-dimensional representation of these successor features (de Cothi and Barry 2020). They test the predictions of the model against environmental manipulations such as dimensional stretches, barrier insertions, and the influence of a room's geometry on the representation of space (related to the behavioural study mentioned above).

Connecting the Past and the Future: Multiscale Predictive Representations

Let us return to the example of planning a flight, and moving between coarse and more detailed scales during planning. In the models described earlier, the scale or horizon of abstraction of a given successor representation is determined by the discount parameter, γ. However, in the agents we discussed so far only one discount parameter was used to learn the relations structure of the environment. Yet, various scales of abstraction might be more appropriate for different planning problems. How does the brain, or an RL agent, handle planning seamlessly at different scales of granularity?

If we take the successor representation approach, one hypothesis is that the brain or the model simply stores multiple successor representations with different values of the discount parameter γ. In other words, the representation of the environment stores multi-step dependencies according to different scales or discount parameters, leading to a multiscale set of successor representations.

If such a model captures multiscale representations in the brain, a number of testable empirical predictions follow. One prediction is that during navigation of long distances, representations in different parts of the brain should reveal sensitivity to different horizons. If not, all brain regions should show the same level of granularity.

In order to test this hypothesis, it is important to go beyond small-scale laboratory studies of how relational knowledge enables inference and planning in few step controlled designs. Such a study needs to be feasible for neural measurements, e.g., human neuroimaging. Virtual reality-like navigation of long distances inside an fMRI scanner offers a sweet spot between laboratory and life-scale planning inside a scanner, offering a window into studying how people use stored knowledge in continuously unfolding navigation, e.g., walking long distances in a city.

One study used an existing fMRI dataset of virtual navigation of realistic distances (of up to kilometres) in the city of Toronto (Brunec et al. 2018). They hypothesized that there are predictive representations organized at multiple scales along posterior–anterior prefrontal and hippocampal hierarchies, and that they guide naturalistic VR navigation (Brunec and Momennejad 2021). To test this hypothesis, the study conducted model-based representational similarity analyses of neuroimaging data measured during navigation of realistically long paths in VR. They tested the pattern similarity of each point along a given path to a weighted sum of its successor points within different predictive horizons (Figure 12.5). They found that the anterior prefrontal cortex (PFC) showed the largest predictive horizons (up to a kilometre), posterior hippocampus the smallest (about 25 metres), with the anterior hippocampus and orbitofrontal regions in between (100–200 metres). These findings offer novel insights into how multiscale cognitive maps can connect memory and prediction to support hierarchical planning (Figure 12.5, bottom).

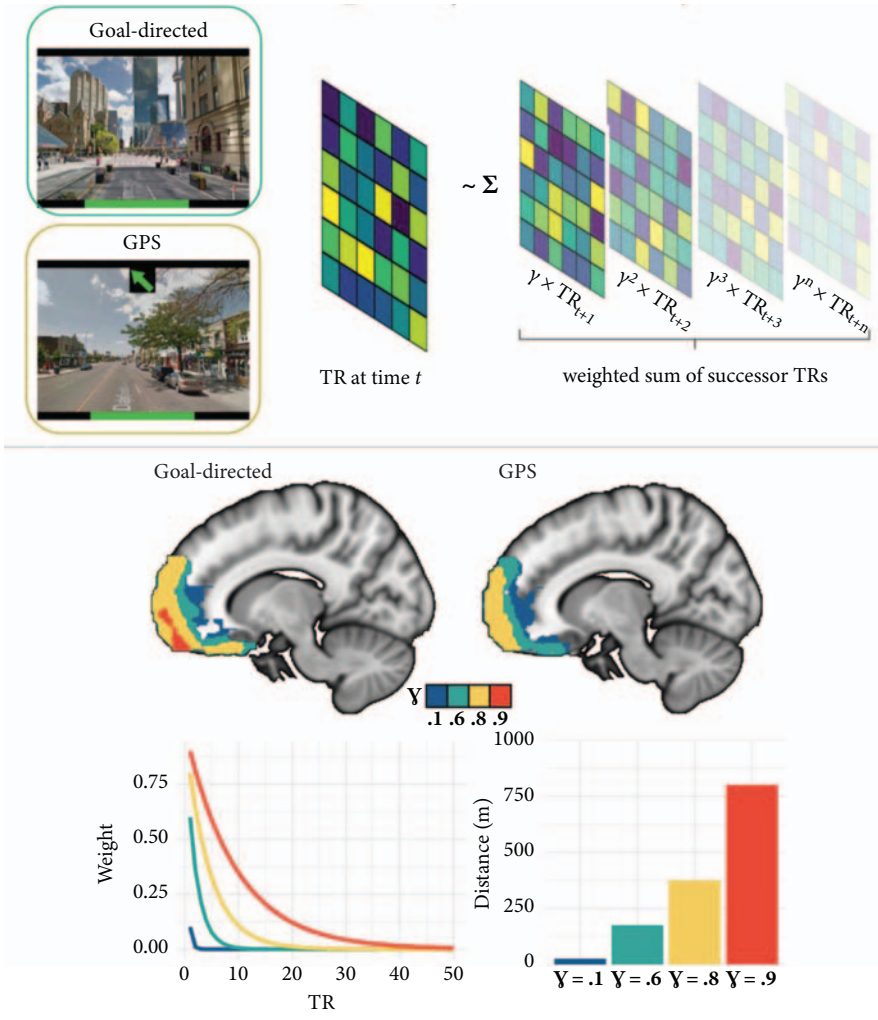


Figure 12.5 Multiscale successor representations. (Top) The virtual Toronto navigation experiment was conducted in fMRI with two conditions: goal-directed, in which participants navigated to a known goal with no guidance and based on their memory; and GPS, in which they did not know the goal and followed a GPS arrow in an unfamiliar part of the city. The similarity of fMRI representations was measured between every point on the trajectory and the weighted sum of its successor states on the taken path using different discount parameters. (Bottom) In the prefrontal cortex, the largest horizon or scale was only observed in the goal-directed condition (in the anterior PFC, up to about 1 kilometre). Smaller horizons, which could have been seen in the VR image, were observed in the GPS condition as well.

A number of recent studies consistent with the multiscale view here are noteworthy. A recent study simultaneously recorded hippocampal-prefrontal ensembles and investigated how rats generalize navigational rules across different environments. They showed that the hippocampus represented the specificity of

separate environments, while the prefrontal cortex representations generalized across environments (Tang, Shin, and Jadhav 2023). A human study investigated abstract representations of a story in human brains as participants listened to the story (Owen, Chang, and Manning 2021). They found that the anterior hippocampus and prefrontal cortex had more information about paragraphs and longer scale themes in the intact stories, than words and jumbled versions of the story.

While some studies propose emergent self-scaling in entorhinal grid fields (Fiete, Khona, and Chandra 2023), other studies explored models of mechanisms underlying multiscale memory retrieval. They suggest that inhibitory diversity can increase the range of memory retrieval and change how the network of memories becomes activated (Burns, Haga 芳賀 達也, and Fukai 深井朋樹 2022). Moreover, recent bat studies have reported multiscale representation of larger spatial environments in the bat hippocampus (Eliav et al. 2021). Taken together, this body of studies offers evidence in support of multiscale memory representations across species and using a diverse set of imaging, empirical, and theoretical methods.

Neural Evidence for Offline Replay

Earlier we discussed the importance of offline replay in the models that best capture human behaviour in flexible planning (Figure 12.3). Let us consider neural evidence that offline replay of past states contributes to updating a planning policy.

A human neuroimaging study used a behavioural design with reward revaluation (where the rewards change but the may of the environment doesn't change, Figure 12.3) and a control condition (Momennejad et al. 2018), and interleaved three 30-second rest periods during the re-learning phase (in which participants no longer visited the starting states). The study showed that while offline replay of earlier states that had not been visited in a while (Figure 12.3, state 1) was correlated with changing planning behaviour in the test phase, there was no correlation between offline replay and behaviour in the control condition (where nothing changes in the world).

Notably, the study did not use hippocampal replay, which is too fast for fMRI and hippocampal fMRI signals are often noisier than some other regions.[3] That said, the reactivation of cortical patterns related to an earlier state. This memory reactivation follows the expected patterns hypothesized by replay that prioritizes surprise. Moreover, the study showed that the extent to which the brain was sensitive to unsigned prediction error (or surprise) during the learning phase, could predict how much the earlier states would be replayed during rest, and in turn, how much this offline replay would correlate with subsequent changes in revaluation behaviour during the test phase.

[3] N.B. Some suggest that specific experimental designs can overcome this problem of fast hippocampal replay (Schuck and Niv 2019). However, it is the author's opinion that the feasibility and reliability of this approach for other study designs that may require longer learning and re-learning phases as well as rest periods remains to be established.

Over the past decade a number of studies have investigated the role of replay in behaviour across different species and using different imaging modalities. A series of human fMRI studies have shown evidence for offline replay in non-spatial sequential tasks across the hippocampus, visual cortex, and the orbitofrontal cortex (Schuck and Niv 2019; Wittkuhn and Schuck 2021; Schuck et al. 2016; Wittkuhn et al. 2021).

A series of rodent studies suggest that forward and backward or reverse replay may differentially contribute to planning and updating a plan following prediction errors, forming a cognitive map by capturing the topological map of the environment (Pfeiffer and Foster 2013; Pfeiffer and Foster 2015b; Foster and Wilson 2006; Wu and Foster 2014; Foster 2017; Ólafsdóttir et al. 2015; Pfeiffer and Foster 2015a; Widloski and Foster 2022). A recent study suggests that hippocampal replay appears after a single experience but episodic details may emerge with more experience (Berners-Lee et al. 2022).

A number of other studies have established the role of prediction errors and surprise (unsigned prediction errors) in memory in both laboratory tasks (Rouhani and Niv 2021) as well as natural and long-term life events (Rouhani, Stanley et al. 2023). Other research has highlighted the role of uncertainty and prediction errors in offline learning and memory (Rouhani, Niv et al. 2023; Schapiro et al. 2018). More research is needed to better understand whether there are varieties of prioritization of memory reactivation and replay depending on the task demands, time pressure, and memory resources (Barnett and Momennejad 2022). It is also possible that such task-specific prioritization is meta-learned during the life-long learning of processes that are optimal for different classes of generalized problems (Wang et al. 2018; Wang et al. 2016; Botvinick et al. 2019).

One area that we did not address in depth within this chapter is the role of sleep in replay and the reorganization of memory. Consistent with models, behavioural, and neural evidence discussed in sections 2 and 3, cognitive neuroscience research suggests that sleep improves memory representations in humans (Coutanche et al. 2013). This body of work lends further evidence to the idea that replay and reorganization of memories during sleep may support generalization, semantic memory, and flexible behaviour (Tandoc et al. 2021; Schapiro et al. 2018; Poe, Walsh, and Bjorness 2010; Schapiro, McDevitt et al. 2017).

The Successor Representation and Episodic Memory

So far we have only discussed hippocampal function in terms of multi-step predictive representations and temporal abstraction. Neuroscience has provided mechanisms and evidence for how synaptic plasticity in the hippocampus leads to sequence learning and spatial maps (Mehta, Lee, and Wilson 2002; Mehta 2015) and at different scales (Moore et al. 2021). However, beyond capturing the relational structure of events, the hippocampus is also involved in episodic memory

(Tulving and Markowitsch 1998), or memory of individual events. Are both these two seemingly different functions supported by successor representations?

A computational study has proposed a link between the successor representation, eligibility traces, and the temporal context model (TCM) of episodic memory (Gershman et al. 2012). According to TCM, as we navigate the world the brain's mental context dynamically changes over time in response to both internal and external events, and anything we store in memory is bound with the temporal context during encoding. Therefore, the temporal context can serve as a cue for retrieval, explaining the way in which memories encoded close in time (i.e., have a shared temporal context) are recalled together as a cluster. TCM offers quantitative explanations of the recency effect (recent memories are easier to recall) and the contiguity effect (adjacent items are easier to recall, especially a successor adjacent item), and related medal temporal lobe function (Howard and Kahana 2002; Polyn, Norman, and Kahana 2009; Howard et al. 2005; Howard, Youker, and Venkatadass 2008; Mau et al. 2018).

The computational study suggested that the Temporal Context Model (TCM) is estimating the SR using temporal difference learning–as described earlier (similar to Equation 3) (Gershman et al. 2012). They suggest that the successor prediction is formed through recurrent dynamics in the CA3 subfield of the hippocampus, which is then compared to sensory input arriving at CA1 directly from the entorhinal cortex (EC), thus, CA1 computes successor prediction error, consistent with the reported novelty (mismatch) signal in this region (Lisman and Otmakhova 2001; Kumaran and Maguire 2005 2007).

A more recent study suggests that error-driven temporal difference learning may not be implemented in hippocampal networks (George et al. 2022). Rather, they suggest that spike-timing dependent plasticity (STDP), a form of Hebbian learning, may rapidly learn an approximation of SR from 'theta sweeps', or temporally compressed trajectories. The model uses spiking neurons modulated by theta-band oscillations to capture SR-related phenomena (e.g., backwards expansion on a 1D track and elongation near walls in 2D (Stachenfeld, Botvinick, and Gershman 2017)), as well as multiscale place field sizes along the dorsal-ventral axis of the rodent hippocampus. The authors suggest that such topological ordering is necessary to prevent larger place fields from mixing up the scales (George et al. 2022).

Another study looked at the mathematical relationship between the successor representation and another prominent model of remembering multiscale memories, which models long term memory using the Laplace transform of past events (Momennejad and Howard 2018). The study noted that SR with a single scale often discards information about the sequential order of states and the distance between them. Given these are task-relevant in many navigation tasks in animals and artificial agents, the paper suggested that establishing the plausibility of SR as an organizing principle for the past and future requires an approach that can reconstruct the sequence.

The authors proposed that operations on an ensemble of SRs with multiple scales can reconstruct both the sequence of future states and estimate the distance to goal (Figure 12.5, bottom left). The computation needed was simply to compute the derivative of SR between a given *S* and *S'*, which can be computed linearly. They showed that a multiscale SR ensemble is mathematically equivalent to the Laplace transform of future states, and the inverse of this Laplace transform is a biologically plausible linear estimation of the derivative. This suggests the possibility that multiscale SR and its derivative could lead to a common principle for how the medial temporal lobe supports both map-based and vector-based navigation.

In short, the study showed that multiscale successor representations and their derivatives (Momennejad and Howard 2018) are mathematically equivalent to a prominent computational model of memory (Shankar and Howard 2012; Shankar, Singh, and Howard 2016), in which the inverse laplace transform of a sequence of past events corresponds to distance of the past memory to the present. Notably, the multiscale SR approach and its derivative could explain distance to goal cells observed in bat hippocampal data (Sarel et al. 2017). Note that not only can individual state trajectories and their distance be recovered from multiscale SR, but since the Laplace transform and its inverse had been previously proposed to underlie remembering past events in a scale-invariant fashion (Shankar and Howard 2012; M. W. Howard et al. 2014; Shankar, Singh, and Howard 2016), these findings add credence to the idea of multiscale SR as a principle for memory organization in the medial temporal lobe.

Predictive Representations and Complementary Learning Systems

A related computational view suggests the complementary learning systems in the hippocampus: that it serves both learning distinct separate episodes as well as general statistical commonalities among the episodes (Schapiro, Turk-Browne et al. 2017). They used a neural network rather than the RL framework to ask how the hippocampus handles both statistical learning (which requires detecting commonalities) and memorization of individual episodes (Schapiro, Turk-Browne et al. 2017). The neural network model, with hippocampus-like connectivity and subfields, was trained on sequences with temporal regularities, similar to the stimuli in statistical learning experiments. The results suggest that the pathway connecting the entorhinal cortex directly to CA1 may support statistical learning (monosynaptic pathway), while the pathway connecting entorhinal cortex to CA1 through dentate gyrus and CA3 (trisynaptic pathway) learns individual episodes.

According to this view, associative reactivation through recurrence may give rise to the representations of statistical regularities, and different anatomical pathways may mediate the trade-off between learning episodes and associative functions of

the hippocampus. This is also in line with a study suggesting distinct roles for dorsal CA3 and CA1 in memory for sequential nonspatial events (Farovik, Dupont, and Eichenbaum 2010). A more recent human neuroimaging study used an associative learning paradigm to test both SR and the complementary learning systems predictions, and found intriguing convergence (Pudhiyidath et al. 2022). Future research can further illuminate this link between SR and complementary learning and memory systems.

Consistently, neuroimaging studies suggest a complementary role for CA1 and CA3 in episodic memory and context. One study showed that CA1 represented objects that shared an episodic context as more similar to each other, while CA3 differentiated between objects encountered in the same episodic context (Dimsdale-Zucker et al. 2018). The authors suggest that the complementary nature of CA1 and CA3 captures how we parse experiences into cohesive episodes while retaining the specific details. A consistent study showed that damage to CA3 can disrupt both recent and distant episodic memories (Miller et al. 2020).

Taken together, these studies suggest that hippocampal subfields may contribute to both associative and episodic learning through complementary circuitry, functions, and interactions. Moreover, recent work connects the statistical learning of successor representation to on-task replay (Wittkuhn, Krippner, and Schuck 2022), providing further credence to the importance of the predictive structure of memory in our understanding of behavioural and neural phenomena surrounding memory, replay, and planning.

Predictive Representations, Hierarchical Planning, and Deep RL

Most real world navigation and planning problems involve hierarchical problem solving. Studies of taxi drivers and indigenous navigators offer evidence for the explicit use of hierarchical planning in humans in the real world (Spiers and Maguire 2008; Fernandez-Velasco and Spiers 2024). Behavioural and neural signatures of hierarchical planning and navigation, as well as their computational models, are active areas of research in contemporary computer science and cognitive neuroscience.

Hierarchical RL (or HRL) focuses on representation learning mechanisms that empower an agent to explore an environment and learn options, which are abstractions of policies or various sequences of state–action–state sequences that functionally achieve the same goal (Sutton, Precup, and Singh 1999; Dietterich n.d.; Stolle and Precup 2002; Barto and Mahadevan 2003; Bacon, Harb, and Precup 2016; Botvinick and Weinstein 2014; Xia and Collins 2020). The HRL and options framework allow for more efficient exploration and representations for goal-directed navigation and problem solving.

Moreover, neuroscientific studies of hierarchical planning and reasoning suggest a key role for the prefrontal cortex, e.g., in motivating functional hierarchies in cognitive control (Egner 2009) and simulating the future (Javadi et al. 2017). Patient studies show that damage to the human anterior prefrontal cortex leads to impairments in multi-tasking, or completing a sequence of errands in order (Burgess 2000; Burgess et al. 2000; Roca et al. 2011). A number of studies combining human fMRI and machine learning approaches have established a role for the anterior prefrontal cortex in prospective memory, or remembering to execute a long-term intention later while we are busy doing something else now (Okuda et al. 2007; Burgess et al. 2008; Momennejad and Haynes 2012; Momennejad and Haynes 2013; Haynes et al. 2015). Primate neural recordings also suggest a role in hierarchical reasoning by the frontal cortex (Sarafyazd and Jazayeri 2019).

Recent work in deep learning used inspiration from human studies of collective memory and collective cognition (Coman et al. 2016; Hirst, Yamashiro, and Coman 2018; Momennejad, Duker, and Coman 2019; Momennejad 2022) to propose a multi-agent approach to hierarchical problem solving using deep RL agents (Nisioti et al. 2022). The study designed a set of hierarchical problems with various levels of branching and breadth, and connected a network of ten deep Q-learning networks (DQNs) with different connectivity graphs, and had each multi-agent collective of agents solve the problems by exploring individually and sharing their replay content with each other (Nisioti et al. 2022). The results revealed that it was not an all-to-all connectivity nor individual agents that could optimally solve all problems. The more difficult problems could only be solved by networks with dynamic connectivity: where agents first communicated in smaller clusters and then changed the connectivity to communicate more broadly, and repeated this dynamic connectivity.

This multi-agent finding is potentially useful in understanding how the prefrontal cortex flexibly coordinates communication among other brain regions to solve different tasks such as hierarchical problems. While some tasks may be better served by dense connectivity among certain brain regions, others may require selective or dynamic connecting of some connections, and the PFC may need to metalearn optimal connectivity patterns for each problem category over the course of continual and lifelong learning (Lopez-Paz and Ranzato 2017; Parisi et al. 2018). Future research could shed light on such a multi-agent communication hypothesis about the role of the PFC in flexible task switching and hierarchical problem solving.

Let us return to multiscale predictive representations. Over the past decade, a series of RL studies have linked the options and hierarchical RL framework to various forms of the successor representation. A number of RL studies have focused on the eigenvectors of the graph Laplacian of the environment's structure to derive options and subgoals in HRL (Anderson and Morley 1985; Gutman 2003; Sprekeler 2011; Machado, Bellemare, and Bowling 2017; Klissarov and Machado 2023), as well as eigen-options, or generalized option components, the linear combination of which can provide various policies in an environment (Sherstan, Machado,

and Pilarski 2018; Machado et al. 2017; Machado, Bellemare, and Bowling 2020). Notably, SR is comparable to the inverse of the graph Laplacian. Thus, a number of studies investigated how successor feature learning and deep successor representation models acquire representational underpinning of HRL.

A standing challenge in this area and for future studies involves the observed dichotomy between the learning rules in deep SR, or deep successor feature agents, and the emergent structure of learned representations. While the learning rules follow SR, sometimes the similarity structure of representations do not reflect what would be expected to a tabular successor representation approach. Future studies are required to better understand efficient and optimal learning and use of SR and deep SR in hierarchical learning and problem solving.

Future research is required to better understand more prefrontal cortical-centred functions such as task sets and schema (Farzanfar et al. 2023; Masís-Obando, Norman, and Baldassano 2022; Preston and Eichenbaum 2013; Graziano and Webb 2015; Gilboa and Marlatte 2017; McKenzie et al. 2014; Gupta et al. 2012), compositional tasks, and decomposing value and policy with different basis sets (Whittington et al. 2019).

Evaluating Generative AI's Planning Behaviour and Underlying Representations

Evaluating Navigation Behaviour in Xbox Games

Contemporary generative AI is ubiquitous, from search engines to medical assistants, office copilots, and game agents, various dimensions of the future of life and work are tied to it (Lee, Goldberg, and Kohane 2023). However, various capacities discussed here related to planning, long-term memory with efficient retrieval, and navigation remain unresolved challenges for contemporary generative AI.

It is the author's belief that cognitive science and computational neuroscience inspired methods and theories can enhance both the evaluation of generative AI systems and inspire architectures and solutions that enhance their abilities. This may herald a new dawn for the application of cognitive neuroscience approaches in AI.

While a detailed survey remains outside the scope of the present chapter, it is noteworthy that the classic computer science benchmarks have proven to not be sufficient for the evaluation of memory, generalization, planning, and navigation, and methods from cognitive and neural science prove highly relevant. For instance, two studies compared the behaviour of human-like agents with human behaviour when navigating in the same VR game environment: Bleeding Edge (Devlin et al. 2021; Zuniga et al. 2022). They showed that while both agents had similarly high performance on benchmarks such as reward and steps to completion, neither of them passed the Navigation Turing Test (HNTT).

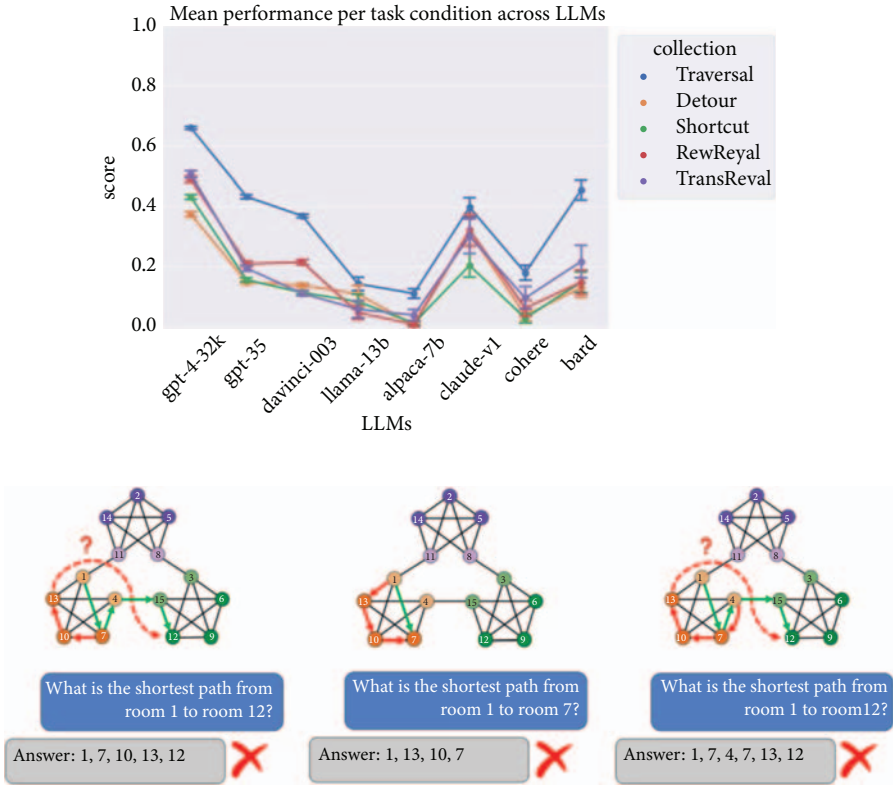


Figure 12.6 Evaluating planning and cognitive maps in large language models. (Top) The performance of eight LLMs on tasks related to flexible planning. (Bottom) Three major failure modes of LLMs include hallucinating edges that do not exist (there is no direct path between 13 and 12), taking a long path when a one-step path is available, and falling in loops (I. Momennejad et al. 2023). These failure modes do not support the idea that LLMs including GPT-4 have emergent cognitive maps or planning ability.

That is, human judges observed side by side videos of an avatar navigating the game, and could significantly tell in which videos the avatar's navigation was controlled by agents and in which by humans (Devlin et al. 2021; Zuniga et al. 2022). Notably, a second study use six different artificial judges to determine if they can judge human-like navigation in the game, and found that while they reached high performance at detecting human play, when comparing videos played by two agents side by side, they could not distinguish or rank which of the two agents would be judged as navigating more human-like by human judges (Devlin et al. 2021).

The significance of these findings are twofold: first, benchmarks are not enough to capture human-like behaviour—see this rubric for human-like agents and neuroAI for more detail (Momennejad 2023); and second, judging which of two agents display human-like navigation was not trivial for existing imitation learning and deep learning models. A deeper understanding of both human navigation behaviour and metrics for its understanding are required to address these challenges.

Can Large Language Models Plan?

Another intersection of the research discussed in this chapter with generative AI is in the evaluation of planning and navigation behaviour in large language models (LLMs) such as GPT-4, Bard, or Llama. A recent study (Momennejad et al. 2023) turned the design of a number of studies mentioned in this paper (including Figure 12.3) into prompts, and tested eight LLM's behaviour on variety of planning related tasks, such as reward or goal directed planning, finding shortest paths during multi-step traversal, reward revaluation, transition revaluation, shortcut, and detour tasks (Figure 12.6).

The paper proposed CogEval (Momennejad et al. 2023), a cognitive-science inspired protocol for the systematic evaluation of cognitive capacities (such as planning and cognitive maps) in LLMs. From Tolman to the present day, flexible planning behaviour in humans and animals has been studied in terms of robustness to various local changes in the environment that have implications for the global policies, which is the same approach used by the researchers to study the ability to extract and use cognitive maps for planning in eight LLMs.

The study found that taken together, none of the eight large language models tested showed consistent flexible behaviour across the tasks. Notably, while larger models (especially GPT-4) showed apparent success on some of the simpler linear planning tasks, a number of failure modes suggest that even GPT-4 does not have strong emergent planning capacities. These failure modes included hallucinating edges and paths that didn't exist, falling into loops, and taking many steps of unnecessary moves even when the LLM merely needed to traverse a one-step path. Together, planning performance and failures in LLMs including GPT-4 do not support the idea of emergent planning or cognitive map capacity.

Follow up research (Webb et al. 2023) inspired by cognitive computational neuroscience methods, improved LLM-based planning by creating iterative prompts that function like different components of the prefrontal cortex (PFC). This led to a modular black-box architecture, where GPT-4 was prompted to play the role of different PFC regions to solve the problem in an iterative fashion (Webb et al. 2023). The results show improvement in hallucinations as well as planning ability in both graph traversal and Tower of Hanoi tasks. Future research can improve on this modular approach by creating multiscale modules that can identify subgoals.

Summary

How do memories simultaneously capture details about the past, while enabling generalization, prediction, and planning? Here we reviewed evidence for the hypothesis that the answer involves how memories are organized in the brain, namely, as multiscale predictive representations. We reviewed computational,

behavioural, and neural evidence for this hypothesis, and showed that it is compatible with complementary memory systems, in which different subregions and gradients of the brain's memory systems collaborate to support episodic details on the one hand, and abstraction on the other: the past and the future.

We discussed reinforcement learning as one of the computational frameworks for understanding how predictive representations may be organized in memory. The successor representation (SR) was discussed as a candidate principle for generalization in computational accounts of memory, and the structure of predictive representations in the hippocampus and the prefrontal cortex.

We discussed how SR is learned when an agent navigates a sequence of states: the SR stores how often, on average, each upcoming state is expected to be visited. This is different from other RL accounts such as model-based RL that learn only one-step associations, in that SR learns expected this future visitation for states that are multi-steps away. A discount or scale parameter determines how many steps into the future SR's generalizations reach, which in turn enables rapid value computation, subgoal discovery (e.g., via SR's eigenvectors), and flexible decision-making in larger decision trees.

When an agent, or the brain, learns multiple SRs with different discount or scale parameters, we have a multiscale set of predictive representations, the rows of which capture the future, or the successors of each state, and the columns of which capture the past, or the predecessors of each state (Figure 12.4). It has been shown that the successor representation (SR) offers a candidate principle for generalization in reinforcement learning (Dayan 1993; Momennejad et al. 2017; Russek et al. 2017) and computational accounts of episodic memory and temporal context (Gershman et al. 2012), with implications for neural representations in the medial temporal lobe (Stachenfeld, Botvinick, and Gershman 2017) and the midbrain dopamine system (Gardner, Schoenbaum, and Gershman 2018).

The central question was echoed in our discussion of complementary learning systems in the medial temporal lobe (Schapiro et al. 2013). The view suggests that hippocampal subfields contribute to both associative and statistical learning, as well as learning of distinct episodes. We discussed how predictive representations and the successor representation could be used to model this theory as well.

We also discussed how a combined computational, behavioural, and neural evaluation methodology can be used to evaluate and improve generative AI. We discussed an example of a deep RL agent navigating an Xbox game, showing that merely beating certain benchmark metrics such as total rewards or steps to goal are not sufficient to guarantee human-like navigation behaviour in the Xbox game. Possible solutions may indeed require a more brain-like approach to the architecture of the agents. We also summarized research evaluating and improving planning behaviour in large language models, using similar empirical paradigms to those summarized in this chapter.

Taken together, evidence reviewed here supports the idea that memories and cognitive maps may be structured in terms of multiscale predictive representations along hippocampal and prefrontal hierarchies, supporting flexible behaviour in

humans, rodents, bats, and agents. Future research can shed further light on the processes that use these representations for further abstraction, schema, and task-based generalization mediated by these regions.

Acknowledgements

The author is profoundly grateful to Kim Stachenfeld for comments on the manuscript, and Lynn Nadel and Sara Aronowitz for their kindness, resourcefulness, and patience in both organizing workshops as well as this book.

References

Bacon, Pierre-Luc, Jean Harb, and Doina Precup. (2016). The Option-Critic Architecture. arXiv:1609.05140 [cs], September. http://arxiv.org/abs/1609.05140

Barnett, Samuel A., and Ida Momennejad. (2022). PARSR: Priority-adjusted replay for successor representations. In *Reinforcement Learning and Decision Making*. https://samuelabarnett.com/papers/PARSR.pdf

Barto, Andrew G., and Sridhar Mahadevan. (2003). Recent advances in hierarchical reinforcement learning. *Discrete Event Dynamic Systems: Theory and Applications* 13(1/2), 41–77.

Behrens, Timothy E. J., Timothy H. Muller, James C. R. Whittington, Shirley Mark, Alon B. Baram, Kimberly L. Stachenfeld, and Zeb Kurth-Nelson. (2018). What is a cognitive map? Organizing knowledge for flexible behavior. *Neuron* 100(2), 490–509.

Bellmund, Jacob L. S., William de Cothi, Tom A. Ruiter, Matthias Nau, Caswell Barry, and Christian F. Doeller. (2020). Deforming the metric of cognitive maps distorts memory. *Nature Human Behaviour*, 4(2), 177–88.

Berners-Lee, Alice, Ting Feng, Delia Silva, Xiaojing Wu, Ellen R. Ambrose, Brad E. Pfeiffer, and David J. Foster. (2022). Hippocampal replays appear after a single experience and incorporate greater detail with more experience. *Neuron*, March. https://doi.org/10.1016/j.neuron.2022.03.010

Botvinick, Matthew, Sam Ritter, Jane X. Wang, Zeb Kurth-Nelson, Charles Blundell, and Demis Hassabis. (2019). Reinforcement learning, fast and slow. *Trends in Cognitive Sciences*, 23(5), 408–22.

Botvinick, Matthew, and Ari Weinstein. (2014). Model-based hierarchical reinforcement learning and human action control. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 369(1655). https://doi.org/10.1098/rstb.2013.0480

Brunec, Iva K., Buddhika Bellana, Jason D. Ozubko, Vincent Man, Jessica Robin, Zhong-Xu Liu, Cheryl Grady et al. (2018). Multiple scales of representation along the hippocampal anteroposterior axis in humans. *Current Biology: CB*, 28(13), 2129–35.e6.

Brunec, Iva K., and Ida Momennejad. (2021). Predictive representations in hippocampal and prefrontal hierarchies. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, November. https://doi.org/10.1523/JNEUROSCI.1327-21.2021

Burgess, Paul W., Iroise Dumontheil, Sam J. Gilbert, Jiro Okuda, Marieke L. Schölvinck, and Jon S. Simons. (2008). On the role of rostral prefrontal cortex (area 10) in prospective memory. In Matthias Kliegel (ed.), *Prospective Memory: Cognitive, Neuroscience, Developmental, and Applied Perspectives* (pp. 235–60). New York, NY: Taylor & Francis Group/Lawrence Erlbaum Associates.

Burgess, P. W. (2000). Strategy application disorder: The role of the frontal lobes in human multitasking. *Psychological Research*, 63(3-4), 279–88.

Burgess, P. W., E. Veitch, A. de Lacy Costello, and T. Shallice. (2000). The cognitive and neuroanatomical correlates of multitasking. *Neuropsychologia*, 38(6), 848–63.

Burns, Thomas F., Tatsuya Haga 芳賀 達也, and Tomoki Fukai 深井朋樹. (2022). Multiscale and extended retrieval of associative memory structures in a cortical model of local-global inhibition balance. *eNeuro*, 9(3). https://doi.org/10.1523/ENEURO.0023-22.2022

Collin, Silvy H. P., Branka Milivojevic, and Christian F. Doeller. (2015). Memory hierarchies map onto the hippocampal long axis in humans. *Nature Neuroscience*, 18(11), 1562–64.

Coman, Alin, Ida Momennejad, Rae D. Drach, and AndraGeana. (2016). Mnemonic convergence in social networks: The emergent properties of cognition at a collective level. *Proceedings of the National Academy of Sciences of the United States of America*, 113(29), 8171–76.

Cothi, William de, and Caswell Barry. (2020). Neurobiological successor features for spatial navigation. *Hippocampus*, 30(12), 1347–55.

Cothi, William de, Nils Nyberg, Eva-Maria Griesbauer, Carole Ghanamé, Fiona Zisch, Julie M. Lefort, Lydia Fletcher et al. (2022). Predictive maps in rats and humans for spatial navigation. *Current Biology: CB*, 32(17), 3676–89.e5.

Coutanche, Marc N., Carol A. Gianessi, Avi J. H. Chanales, Kate W. Willison, and Sharon L. Thompson-Schill. (2013). The role of sleep in forming a memory representation of a two-dimensional space. *Hippocampus*, 23(12), 1189–97.

Daw, Nathaniel D., and Peter Dayan. (2014). The algorithmic anatomy of model-based evaluation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 369(1655). https://doi.org/10.1098/rstb.2013.0478

Daw, Nathaniel D., Samuel J. Gershman, Ben Seymour, Peter Dayan, and Raymond J. Dolan. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, 69(6), 1204–15.

Daw, Nathaniel D., Yael Niv, and Peter Dayan. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, 8(12), 1704–11.

Dayan, Peter. (1993. Improving generalization for temporal difference learning: the successor representation. *Neural Computation*, 5(4), 613–24.

Devlin, Sam, Raluca Georgescu, Ida Momennejad, Jaroslaw Rzepecki, Evelyn Zuniga, Gavin Costello, Guy Leroy, Ali Shaw, and Katja Hofmann. (2021). Navigation Turing Test (NTT), learning to evaluate human-like navigation. *ICML*, May. http://arxiv.org/abs/2105.09637

Dietterich, Thomas G. (n.d.). The MAXQ method for hierarchical reinforcement learning. https://axon.cs.byu.edu/Dan/778/papers/Hierarchical%20Reinforcement%20Learning/Dietterich1.pdf (accessed 25 July 2023).

Dimsdale-Zucker, Halle R., Maureen Ritchey, Arne D. Ekstrom, Andrew P. Yonelinas, and Charan Ranganath. (2018). CA1 and CA3 differentially support spontaneous retrieval of episodic contexts within human hippocampal subfields. *Nature Communications*, 9(1), 294.

Egner, Tobias. (2009). Prefrontal cortex and cognitive control: Motivating functional hierarchies. *Nature Neuroscience*.

Eliav, Tamir, Shir R. Maimon, Johnatan Aljadeff, Misha Tsodyks, Gily Ginosar, Liora Las, and Nachum Ulanovsky. (2021). Multiscale representation of very large environments in the hippocampus of flying bats. *Science*, 372(6545). https://doi.org/10.1126/science.abg4020

Epstein, Russell A., Eva ZIta Patai, Joshua B. Julian, and Hugo J. Spiers. (2017). The cognitive map in humans: Spatial navigation and beyond. *Nature Neuroscience*, 20(11), 1504–13.

Farovik, Anja, Laura M. Dupont, and Howard Eichenbaum. (2010). Distinct roles for dorsal CA3 and CA1 in memory for sequential nonspatial events. *Learning & Memory*, 17(1), 12–17.

Farzanfar, Delaram, Hugo J. Spiers, Morris Moscovitch, and R. Shayna Rosenbaum. (2023). From cognitive maps to spatial schemas. *Nature Reviews. Neuroscience*, 24(2), 63–79.

Fernandez-Velasco, Pablo, and Hugo J. Spiers. (2024). Wayfinding across ocean and tundra: What traditional cultures teach us about navigation. *Trends in Cognitive Sciences*, 28(1), 56–71.

Fiete, Ila, Mikail Khona, and Sarthak Chandra. (2023). Emergence of robust global modules from local interactions and smooth gradients. *Research Square*. https://doi.org/10.21203/rs.3.rs-2929056/v1

Foster, David J. (2017). Replay comes of age. *Annual Review of Neuroscience*, 40 (July), 581–602.

Foster, David J., and Matthew A. Wilson. (2006). Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature*, 440(7084), 680–83.

Fyhn, Marianne, Sturla Molden, Menno P. Witter, Edvard I. Moser, and May-Britt Moser. (2004). Spatial representation in the entorhinal cortex. *Science*, 305(5688), 1258–64.

Garvert, Mona M., Raymond J. Dolan, and Timothy Ej Behrens. (2017). A map of abstract relational knowledge in the human hippocampal-entorhinal cortex. *eLife*, 6. https://doi.org/10.7554/eLife.17086

Geerts, Jesse P., Fabian Chersi, Kimberly L. Stachenfeld, and Neil Burgess. (2020). A general model of hippocampal and dorsal striatal learning and decision making. *Proceedings of the National Academy of Sciences of the United States of America*, 117(49), 31427–37.

Geerts, J. P., K. L. Stachenfeld, and N. Burgess. (2019). Probabilistic successor representations with kalman temporal differences. *arXiv Preprint*. arXiv:1910.02532. http://arxiv.org/abs/1910.02532

George, Tom M., William de Cothi, Kimberly Stachenfeld, and Caswell Barry. (2022). Rapid learning of predictive maps with STDP and theta phase precession. *bioRxiv*. https://doi.org/10.1101/2022.04.20.488882

Gershman, Samuel J., Christopher D. Moore, Michael T. Todd, Kenneth A. Norman, and Per B. Sederberg. (2012). The successor representation and temporal context. *Neural Computation*, 24(6), 1553–68.

Gilboa, Asaf, and Hannah Marlatte. (2017). Neurobiology of schemas and schema-mediated memory. *Trends in Cognitive Sciences*, 21(8), 618–31.

Gläscher, Jan, Nathaniel Daw, Peter Dayan, and John P. O'Doherty. (2010). States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, 66(4), 585–95.

Graziano, Michael S. A., and Taylor W. Webb. (2015). The attention schema theory: A mechanistic account of subjective awareness. *Frontiers in Psychology*, 6 (April), 500.

Gupta, Anoopum S., Matthijs A. A. van der Meer, David S. Touretzky, and A. David Redish. (2012). Segmentation of spatial experience by hippocampal theta sequences. *Nature Neuroscience*, 15(7), 1032–39.

Gutman, I. (2003). Some properties of Laplacian eigenvectors. *Bulletin*, 28, 1–6.

Hafner, Danijar, Timothy Lillicrap, Mohammad Norouzi, and Jimmy Ba. (2020). Mastering Atari with discrete world models. *arXiv [cs.LG]*. arXiv. http://arxiv.org/abs/2010.02193

Hafting, Torkel, Marianne Fyhn, Sturla Molden, May-Britt Moser, and Edvard I. Moser. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, 436(7052), 801–06.

Haynes, J. D., D. Wisniewski, K. Görgen, I. Momennejad, and C. Reverberi. (2015). FMRI decoding of intentions: Compositionality, hierarchy and prospective memory. In *The 3rd International Winter Conference on Brain-Computer Interface*, 1–3.

Hirst, William, Jeremy K. Yamashiro, and Alin Coman. (2018). Collective memory from a psychological perspective. *Trends in Cognitive Sciences*, 22(5), 438–51.

Howard, Lorelei R., Amir Homayoun Javadi, Yichao Yu, Ravi D. Mill, Laura C. Morrison, Rebecca Knight, Michelle M. Loftus, Laura Staskute, and Hugo J. Spiers. (2014). The hippocampus and entorhinal cortex encode the path and euclidean distances to goals during navigation. *Current Biology: CB*, 24(12), 1331–40.

Howard, Marc W., Mrigankka S. Fotedar, Aditya V. Datey, and Michael E. Hasselmo. (2005). The temporal context model in spatial navigation and relational learning: Toward a common explanation of medial temporal lobe function across domains. *Psychological Review*, 112(1), 75–116.

Howard, Marc W., and Michael J. Kahana. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, 46(3), 269–99.

Howard, Marc W., Christopher J. MacDonald, Zoran Tiganj, Karthik H. Shankar, QiAn Du, Michael E. Hasselmo, and Howard Eichenbaum. (2014). A unified mathematical framework for coding time, space, and sequences in the hippocampal region. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 34(13), 4692–707.

Howard, Marc W., Tess E. Youker, and Vijay S. Venkatadass. (2008). The persistence of memory: Contiguity effects across hundreds of seconds. *Psychonomic Bulletin & Review*, 15(1), 58–63.

Javadi, Amir-Homayoun, Beatrix Emo, Lorelei Howard, Fiona Zisch, Yichao Yu, Rebecca Knight, Joao Pinelo Silva, and Hugo Spiers. (2017). Hippocampal and prefrontal processing of network topology to simulate the future. *Nature Communications*, 8 (March), 14652.

Jr, William N. Anderson, and Thomas D. Morley. (1985). Eigenvalues of the Laplacian of a graph. *Linear and Multilinear Algebra*, 18(2), 141–45.

Klissarov, Martin, and Marlos Machado. (2023). Deep Laplacian-based options for temporally-extended exploration. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning* (202:17198–217). Proceedings of Machine Learning Research. PMLR.

Kumaran, Dharshan, and Eleanor A. Maguire. (2005). The human hippocampus: Cognitive maps or relational memory? *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 25(31), 7254–59.

Kumaran, Dharshan, and Eleanor A. Maguire. (2007). Match–mismatch processes underlie human hippocampal responses to associative novelty. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 27(32), 8517–24.

Lee, Peter, Carey Goldberg, and Isaac Kohane. (2023). *The AI Revolution in Medicine: GPT-4 and Beyond*. Pearson.

Lisman, J. E., and N. A. Otmakhova. (2001). Storage, recall, and novelty detection of sequences by the hippocampus: Elaborating on the SOCRATIC model to account for normal and aberrant effects of dopamine. *Hippocampus*, 11(5), 551–68.

Lopez-Paz, David, and Marc' aurelio Ranzato. (2017). Gradient Episodic Memory for Continual Learning. *arXiv [cs.LG]*. arXiv. http://arxiv.org/abs/1706.08840

Machado, Marlos C., Andre Barreto, Doina Precup, and Michael Bowling. (2023). Temporal abstraction in reinforcement learning with the successor representation. *Journal of Machine Learning Research: JMLR*, 24(80), 1–69.

Machado, Marlos C., Marc G. Bellemare, and Michael Bowling. (2017). A Laplacian framework for option discovery in reinforcement learning. In Proceedings of the 34th International Conference on Machine Learning: Volume 70 (pp. 2295–304). ICML'17. JMLR.org.

Machado, Marlos C., Marc G. Bellemare, and Michael Bowling. (2020). Count-based exploration with the successor representation. *Proceedings of the AAAI Conference on Artificial Intelligence* 34(04), 5125–33.

Machado, Marlos C., Clemens Rosenbaum, Xiaoxiao Guo, Miao Liu, Gerald Tesauro, and Murray Campbell. (2017). Eigenoption discovery through the deep successor representation. https://www.arxiv-vanity.com/papers/1710.11089/

Masís-Obando, Rolando, Kenneth A. Norman, and Christopher Baldassano. (2022). Schema representations in distinct brain networks support narrative memory during encoding and retrieval. *eLife*, 11 (April). https://doi.org/10.7554/eLife.70445

Mattar, Marcelo G., and Nathaniel D. Daw. (2018). Prioritized memory access explains planning and hippocampal replay. *Nature Neuroscience*, 21(11), 1609–17.

Mau, William, David W. Sullivan, Nathaniel R. Kinsky, Michael E. Hasselmo, Marc W. Howard, and Howard Eichenbaum. (2018). The same hippocampal CA1 population simultaneously codes temporal information over multiple timescales. *Current Biology: CB*, 28(10), 1499–1508.e4.

McKenzie, Sam, Andrea J. Frank, Nathaniel R. Kinsky, Blake Porter, Pamela D. Rivière, and Howard Eichenbaum. (2014). Hippocampal representation of related and opposing memories develop within distinct, hierarchically organized neural schemas. *Neuron*, 83(1), 202–15.

Mehta, Mayank R. (2015). From synaptic plasticity to spatial maps and sequence learning. *Hippocampus*, 25(6), 756–62.

Mehta, Mayank R., Carol A. Barnes, and Bruce L. McNaughton. (1997). Experience-dependent, asymmetric expansion of hippocampal place fields. *Proceedings of the National Academy of Sciences*, 94(16), 8918–21.

Mehta, M. R., A. K. Lee, and M. A. Wilson. (2002). Role of experience and oscillations in transforming a rate code into a temporal code. *Nature*, 417(6890), 741–46.

Mehta, M. R., M. C. Quirk, and M. A. Wilson. (2000). Experience-dependent asymmetric shape of hippocampal receptive fields. *Neuron*, 25(3), 707–15.

Miller, Thomas D., Trevor T-J Chong, Anne M. Aimola Davies, Michael R. Johnson, Sarosh R. Irani, Masud Husain, Tammy Wc Ng et al. (2020). Human hippocampal CA3 damage disrupts both recent and remote episodic memories. *eLife*, 9 (January). https://doi.org/10.7554/eLife.41836

Momennejad, Ida. (2020). Learning structures: Predictive representations, replay, and generalization. *Current Opinion in Behavioral Sciences*, 32 (April), 155–66.

Momennejad, Ida. (2022). Collective minds: Social network topology shapes collective cognition. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 377(1843). 20200315.

Momennejad, Ida. (2023). A rubric for human-like agents and NeuroAI. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 378(1869). 20210446.

Momennejad, Ida, Ajua Duker, and Alin Coman. (2019). Bridge ties bind collective memories. *Nature Communications*, 10(1), 1578.

Momennejad, Ida, and John-Dylan Haynes. (2013). Encoding of prospective tasks in the human prefrontal cortex under varying task loads. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 33(44), 17342–49.

Momennejad, I., H. Hasanbeig, F. Vieira Frujeri, H. Sharma, R. Ness, H. Palangi, and J. Larson. (2023). Evaluating cognitive maps in large language models with cogeval: no emergent planning. *Advances in Neural Information Processing Systems*.

Momennejad, I., and J. D. Haynes. (2012). Human anterior prefrontal cortex encodes the 'what' and 'when' of future intentions. *NeuroImage*. https://www.sciencedirect.com/science/article/pii/S1053811912002649

Momennejad, I., and M. W. Howard. (2018). Predicting the future with multi-scale successor representations. *bioRxiv*. https://www.biorxiv.org/content/10.1101/449470v1.abstract

Momennejad, I., A. R. Otto, N. D. Daw, and K. A. Norman. (2018). Offline replay supports planning in human reinforcement learning. eLife. https://elifesciences.org/articles/32548

Momennejad, I., E. M. Russek, J. H. Cheong, M. M. Botvinick, N. D. Daw, and S. J. Gershman. (2017). The successor representation in human reinforcement learning. *Nature Human Behaviour*, 1(9), 680–92.

Moore, Andrew W., and Christopher G. Atkeson. (1993). Prioritized sweeping: Reinforcement learning with less data and less time. *Machine Learning*, 13(1), 103–30.

Moore, Jason J., Jesse D. Cushman, Lavanya Acharya, Briana Popeney, and Mayank R. Mehta. (2021). Linking hippocampal multiplexed tuning, hebbian plasticity and navigation. *Nature*, 599(7885), 442–48.

Moser, Edvard I., Emilio Kropff, and May-Britt Moser. (2008). Place cells, grid cells, and the brain's spatial representation system. *Annual Review of Neuroscience*, 31, 69–89.

Nielson, Dylan M., Troy A. Smith, Vishnu Sreekumar, Simon Dennis, and Per B. Sederberg. (2015). Human hippocampus represents space and time during retrieval of real-world memories. *Proceedings of the National Academy of Sciences of the United States of America*, 112(35), 11078–83.

Nisioti, Eleni, Mateo Mahaut, Pierre-Yves Oudeyer, Ida Momennejad, and Clément Moulin-Frier. (2022). Social network structure shapes innovation: experience-sharing in RL with SAPIENS. *arXiv [cs.AI]*. arXiv. http://arxiv.org/abs/2206.05060

O'Keefe, J. (1976). Place units in the hippocampus of the freely moving rat. *Experimental Neurology*, 51(1), 78–109.

O'Keefe, J., and J. Dostrovsky. (1971). The hippocampus as a spatial map. preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, 34(1), 171–75.

O'Keefe, John, and Lynn Nadel. (1978). *The Hippocampus as a Cognitive Map*. Oxford: Clarendon Press.

Okuda, Jiro, Toshikatsu Fujii, Hiroya Ohtake, Takashi Tsukiura, AtsusHi Yamadori, Christopher D. Frith, and Paul W. Burgess. (2007). Differential involvement of regions of rostral prefrontal cortex (Brodmann area 10) in time- and event-based prospective memory. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, 64(3), 233–46.

Ólafsdóttir, H. Freyja, Caswell Barry, Aman B. Saleem, Demis Hassabis, and Hugo J. Spiers. (2015). Hippocampal place cells construct reward related sequences through unexplored space. *eLife*, 4. https://doi.org/10.7554/eLife.06063

Owen, Lucy L. W., Thomas H. Chang, and Jeremy R. Manning. (2021). High-level cognition during story listening is reflected in high-order dynamic correlations in neural activity patterns. *Nature Communications*, 12(1), 5728.

Parisi, German I., Ronald Kemker, Jose L. Part, Christopher Kanan, and Stefan Wermter. (2018). Continual lifelong learning with neural networks: A review. arXiv [cs.LG]. arXiv. http://arxiv.org/abs/1802.07569

Patai, Eva Zita, and Hugo J. Spiers. (2021). The versatile wayfinder: Prefrontal contributions to spatial navigation. *Trends in Cognitive Sciences*, 25(6), 520–33.

Pfeiffer, B. E., and D. J. Foster. (2015a). Autoassociative dynamics in the generation of sequences of hippocampal place cells. *Science*. https://science.sciencemag.org/content/349/6244/180.short

Pfeiffer, B. E., and D. J. Foster. (2015b). Discovering the brain's cognitive map. *JAMA Neurology*. https://jamanetwork.com/journals/jamaneurology/article-abstract/2088875

Pfeiffer, Brad E., and David J. Foster. (2013). Hippocampal place-cell sequences depict future paths to remembered goals. *Nature*, 497(7447), 74–79.

Poe, Gina R., Christine M. Walsh, and Theresa E. Bjorness. (2010). Cognitive neuroscience of sleep. *Progress in Brain Research*, 185: 1–19.

Polyn, Sean M., Kenneth A. Norman, and Michael J. Kahana. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, 116(1), 129–56.

Preston, Alison R., and Howard Eichenbaum. (2013). Interplay of hippocampus and prefrontal cortex in memory. *Current Biology: CB*, 23(17), R764–73.

Pudhiyidath, Athula, Neal W. Morton, Rodrigo ViverOs Duran, Anna C. Schapiro, Ida Momennejad, Demitrius M. Hinojosa-Rowland, Robert J. Molitor, and Alison R. Preston. (2022). Representations of temporal community structure in hippocampus and precuneus predict inductive reasoning decisions. *Journal of Cognitive Neuroscience*, 34(10), 1736–60.

Roca, María, Teresa Torralva, Ezequiel Gleichgerrcht, Alexandra Woolgar, Russell Thompson, John Duncan, and Facundo Manes. (2011). The role of area 10 (BA10) in human multitasking and in social cognition: A lesion study. *Neuropsychologia*, 49(13), 3525–31.

Rouhani, Nina, and Yael Niv. (2021). Signed and unsigned reward prediction errors dynamically enhance learning and memory. *eLife*, 10 (March). e61077.

Rouhani, Nina, Yael Niv, Michael J. Frank, and Lars Schwabe. (2023). Multiple routes to enhanced memory for emotionally relevant events. *Trends in Cognitive Sciences*, July. https://doi.org/10.1016/j.tics.2023.06.006

Rouhani, Nina, Damian Stanley, COVID-Dynamic Team, and Ralph Adolphs. (2023). Collective events and individual affect shape autobiographical memory. *Proceedings of the National Academy of Sciences of the United States of America*, 120(29). e2221919120.

Russek, Evan M., Ida Momennejad, Matthew M. Botvinick, Samuel J. Gershman, and Nathaniel D. Daw. (2017). Predictive representations can link model-based reinforcement learning to model-free mechanisms. *PLoS Computational Biology*, 13(9). e1005768.

Russek, Evan M., Ida Momennejad, Matthew M. Botvinick, Samuel J. Gershman, and Nathaniel D. Daw. (2021). Neural evidence for the successor representation in choice evaluation. *bioRxiv*. https://doi.org/10.1101/2021.08.29.458114

Samborska, Veronika, James L. Butler, Mark E. Walton, Timothy E. J. Behrens, and Thomas Akam. (2022). Complementary task representations in hippocampus and prefrontal cortex for generalizing the structure of problems. *Nature Neuroscience*, 25(10), 1314–26.

Sarafyazd, Morteza, and Mehrdad Jazayeri. (2019). Hierarchical reasoning by neural circuits in the frontal cortex. *Science*, 364(6441). https://doi.org/10.1126/science.aav8911

Sarel, Ayelet, Arseny Finkelstein, Liora Las, and Nachum Ulanovsky. (2017). Vectorial representation of spatial goals in the hippocampus of bats. *Science*, 355(6321), 176–80.

Schapiro, Anna C., Elizabeth A. McDevitt, Lang Chen, Kenneth A. Norman, Sara C. Mednick, and Timothy T. Rogers. (2017). Sleep benefits memory for semantic category structure while preserving exemplar-specific information. *Scientific Reports*, 7(1), 14869.

Schapiro, Anna C., Elizabeth A. McDevitt, Timothy T. Rogers, Sara C. Mednick, and Kenneth A. Norman. (2018). Human hippocampal replay during rest prioritizes weakly learned information and predicts memory performance. *Nature Communications*, 9(1), 3920.

Schapiro, Anna C., Timothy T. Rogers, Natalia I. Cordova, Nicholas B. Turk-Browne, and Matthew M. Botvinick. (2013). Neural representations of events arise from temporal community structure. *Nature Neuroscience*, 16(4), 486–92.

Schapiro, Anna C., Nicholas B. Turk-Browne, Matthew M. Botvinick, and Kenneth A. Norman. (2017). Complementary learning systems within the hippocampus: A neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 372(1711). https://doi.org/10.1098/rstb.2016.0049

Schaul, Tom, John Quan, Ioannis Antonoglou, and David Silver. (2015). Prioritized experience replay. *arXiv [cs.LG]*. arXiv. http://arxiv.org/abs/1511.05952

Schuck, Nicolas W., Ming Bo Cai, Robert C. Wilson, and yael niv. (2016). human orbitofrontal Cortex Represents a Cognitive Map of State Space. *Neuron*, 91(6), 1402–12.

Schuck, Nicolas W., and Yael Niv. (2019). Sequential replay of nonspatial task states in the human hippocampus. *Science*, 364(6447). https://doi.org/10.1126/science.aaw5181

Shankar, Karthik H., and Marc W. Howard. (2012). A scale-invariant internal representation of time. *Neural Computation*, 24(1), 134–93.

Shankar, Karthik H., Inder Singh, and Marc W. Howard. (2016). Neural mechanism to simulate a scale-invariant future. *Neural Computation*, 28(12), 2594–627.

Sherstan, Craig, Marlos C. Machado, and Patrick M. Pilarski. (2018). Accelerating learning in constructive predictive frameworks with the successor representation. *arXiv [cs.LG]*. arXiv. http://arxiv.org/abs/1803.09001

Spiers, Hugo J., and Eleanor A. Maguire. (2007). A navigational guidance system in the human brain. *Hippocampus*, 17(8), 618–26.

Spiers, Hugo J., and Eleanor A. Maguire. (2008). The dynamic nature of cognition during wayfinding. *Journal of Environmental Psychology*, 28(3), 232–49.

Sprekeler, Henning. (2011). On the relation of slow feature analysis and laplacian eigenmaps. *Neural Computation*, 23(12), 3287–302.

Stachenfeld, Kimberly L., Matthew M. Botvinick, and Samuel J. Gershman. (2017). The hippocampus as a predictive map. *Nature Neuroscience*, 20(11), 1643–53.

Stolle, Martin, and Doina Precup. (2002). Learning options in reinforcement learning. In Sven Koenig and Robert C. Holte (eds.), *Abstraction, Reformulation, and Approximation* (pp. 212–23). Lecture Notes in Computer Science. Springer Berlin Heidelberg.

Strange, Bryan A., Menno P. Witter, Lein, Ed S., and Moser, Edvard I. (2014). Functional organization of the hippocampal longitudinal axis. *Nature Reviews. Neuroscience*, 15(10), 655–69.

Sutton, Richard S. (1991). Dyna, an integrated architecture for learning, planning, and reacting. *SIGART Bulletin*, 2(4), 160–63.

Sutton, Richard S., and Andrew G. Barto. (2018). *Reinforcement Learning: An Introduction*. MIT Press.

Sutton, Richard S., Marlos C. Machado, G. ZachArias Holland, David Szepesvari, Finbarr Timbers, Brian Tanner, and Adam White. (2023). Reward-respecting subtasks for model-based reinforcement learning. *Artificial Intelligence*, 324 (November), 104001.

Sutton, Richard S., Doina Precup, and Satinder Singh. (1999). Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, 112(1), 181–211.

Sutton, Richard S., Csaba Szepesvari, Alborz Geramifard, and Michael P. Bowling. (2012). Dyna-style planning with linear function approximation and prioritized sweeping. arXiv:1206.3285 [cs], June. http://arxiv.org/abs/1206.3285

Tandoc, Marlie C., Mollie Bayda, Craig Poskanzer, Eileen Cho, Roy Cox, Robert Stickgold, and Anna C. Schapiro. (2021). Examining the effects of time of day and sleep on generalization. *PLOS ONE*, 16(8). e0255423.

Tang, Wenbo, Justin D. Shin, and Shantanu P. Jadhav. (2023). Geometric transformation of cognitive maps for generalization across hippocampal-prefrontal circuits. *Cell Reports*, 42(3), 112246.

Tarder-Stoll, Hannah, Christopher Baldassano, and Mariam Aly. (2023). The brain hierarchically represents the past and future during multistep anticipation. *bioRxiv*. https://doi.org/10.1101/2023.07.24.550399

The Nobel Prize in Physiology or Medicine 2014. (2014). NobelPrize.org. https://www.nobelprize.org/prizes/medicine/2014/press-release/ (accessed 10 May 2023).

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, 55(4), 189–208.

Tulving, E., and H. J. Markowitsch. (1998). Episodic and declarative memory: Role of the hippocampus. *Hippocampus*, 8(3), 198–204.

Wang, Jane X., Zeb Kurth-Nelson, Dharshan Kumaran, Dhruva Tirumala, Hubert Soyer, Joel Z. Leibo, Demis Hassabis, and Matthew Botvinick. (2018). Prefrontal cortex as a meta-reinforcement learning system. *Nature Neuroscience*, 21(6), 860–68.

Wang, J. X., Z. Kurth-Nelson, D. Tirumala, and H. Soyer. (2016). Learning to reinforcement learn. *arXiv Preprint arXiv*. https://arxiv.org/abs/1611.05763

Wan, Yi, Ali Rahimi-Kalahroudi, Janarthanan Rajendran, Ida Momennejad, Sarath Chandar, and Harm van Seijen. (2022). Towards evaluating adaptivity of model-based reinforcement learning methods. In *Proceedings of the International Conference on Machine Learning (ICML)*. http://arxiv.org/abs/2204.11464.

Webb, Taylor, Shanka Subhra Mondal, Chi Wang, Brian Krabach, and Ida Momennejad. (2023). A prefrontal cortex-inspired architecture for planning in large language models. *arXiv [cs.AI]*. arXiv. http://arxiv.org/abs/2310.00194

Whittington, James C. R., Timothy H. Muller, Shirley Mark, Guifen Chen, Caswell Barry, Neil Burgess, and Timothy E. J. Behrens. (2019). The Tolman-Eichenbaum machine: Unifying space and relational memory through generalisation in the hippocampal formation. *bioRxiv*. https://doi.org/10.1101/770495

Widloski, John, and David J. Foster. (2022). Flexible rerouting of hippocampal replay sequences around changing barriers in the absence of global place field remapping. *Neuron*, 110(9), 1547–58.e8.

Wittkuhn, Lennart, Samson Chien, Sam Hall-mcmaster, and Nicolas W. Schuck. (2021). Replay in minds and machines. *Neuroscience and Biobehavioral Reviews*, 129 (October), 367–88.

Wittkuhn, Lennart, Lena M. Krippner, and Nicolas W. Schuck. (2022). Statistical learning of successor representations is related to on-task replay. *bioRxiv*. https://doi.org/10.1101/2022.02.02.478787

Wittkuhn, Lennart, and Nicolas W. Schuck. (2021). Dynamics of fMRI patterns reflect sub-second activation sequences and reveal replay in human visual cortex. *Nature Communications*, 12(1), 1–22.

Wu, Xiaojing, and David J. Foster. (2014). Hippocampal replay captures the unique topological structure of a novel environment. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 34(19), 6459–69.

Xia, Liyu, and Anne Gabrielle Eva Collins. (2020). Temporal and state abstractions for efficient learning, transfer and composition in humans. *bioRxiv*. https://doi.org/10.1101/2020.02.20.958587

Zuniga, Evelyn, Stephanie Milani, Guy Leroy, Jaroslaw Rzepecki, Raluca Georgescu, Ida Momennejad, Dave Bignell et al. (2022). How humans perceive human-like behavior in video game navigation. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*, 1–11. CHI EA' 22 391. New York, NY, USA: Association for Computing Machinery.

Index

For the benefit of digital users, indexed terms that span two pages (e.g., 52–53) may, on occasion, appear on only one of those pages.