REMEMBERING WITHOUT STORING: BEYOND ARCHIVAL MODELS IN THE SCIENCE AND PHILOSOPHY OF HUMAN MEMORY

by

Ian O'Loughlin

A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Philosophy
in the Graduate College of
The University of Iowa

August 2014

Thesis Supervisor: Professor David Stern

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Ian O'Loughlin

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy
degree in Philosophy at the August 2014 graduation.

Thesis Committee: ___________________________________
David Stern, Thesis Supervisor

____
Carrie Figdor

____
Gregory Landini

____
Katarina Perovic

____
Richard Fumerton

It's very hard to stop doing things you're used to doing. You almost have to dismantle yourself, scatter it all around, and then put a blindfold on and put it back together so that you avoid old habits.

Tom Waits, *The Beat Goes On*

ACKNOWLEDGEMENTS

There are many people who have played instrumental roles in the development of this project. The following is not intended to be an exhaustive list of these, but rather to highlight and credit several parties who deserve special mention.

I would like to extend heartfelt gratitude to my thesis supervisor, David Stern. David is an excellent teacher, an inspiring scholar, and an exemplary advisor. This project has roots in my experience in his Wittgenstein course during my first semester at Iowa, where I benefitted from David's acute sense of intellectual subtlety and careful attention to differing perspectives, and it has depended on his continued insight and encouragement since. A friend and colleague once told me that we don't necessarily adopt the views of our advisors, but we do inherit their habits of mind and approaches to philosophy. I would be thrilled to inherit these from my advisor.

Along with the study of Wittgenstein, language, and artificial intelligence with David, this project has also been deeply informed by studying cognitive science, complex systems, and mind with Carrie Figdor. Carrie's impressively broad grasp of cognitive science—a wide-ranging, multidisciplinary, and rapidly developing field—has been inspiring and helpful. The project has also benefitted by studies of Aristotelian metaphysics with James Duerlinger, the philosophy of science with Evan Fales, and early modern theories of mind with David Cunning (who has also been a generous graduate director throughout my time in the Ph.D. program at Iowa). I would also like to thank the rest of my thesis committee, Katarina Perovic, Gregory Landini, and Richard Fumerton--not just for their help with this project and engaging discussions, but also for making the Iowa department the sort of place where good philosophy happens.

During the final year of my dissertation, I have been supported by a Ballard-Seashore fellowship through the graduate college, whose staff have graciously accommodated administrative anomalies. I have also been supported by assorted fellowships through the philosophy department at Iowa and through the University's International Programs office, enabling me to pursue research at home and abroad. Without such support this project would not exist.

Perhaps most importantly, I want to thank all of my fellow graduate students in the Iowa philosophy department, and several other philosophically-inclined friends with whom I have been constantly engaged in interweaved discussions of philosophical traditions, mind, science, knowledge, and human nature for the last several years. Among many others, I have been privileged to enjoy the discourse and company of Nathanial Blower, Abraham Graber, Kathryn Noack, Tap Rossetto, Tyler Sproule, and Samuel Taylor, each of whom contributed significantly to the development of the ideas in this project. Many more contributed to my development as a philosopher.

Lastly, I want to thank my partner, Kristin, and my daughter, Fiona, for putting up with late hours, disheveled piles of books, and intermittent rants about protein synthesis inhibitors or the semantics of neural network activation patterns for the last two years. Researching, developing, and writing this dissertation has been a formidable process, and it could not have succeeded without the grace and encouragement of those close to me.

TABLE OF CONTENTS

CHAPTER

I. MEMORY AND WITTGENSTEIN 1

1.1 Memory in philosophy and cognitive science 5
1.2 Wittgenstein and the philosophy of mind and cognitive science 12
1.3 New directions in cognitive science 16

II. COMPUTATIONALISM AND MEMORY 22

2.1 Computationalism and its legacy 23
2.2 Modeling memory 32
2.3 Computationalist assumptions in traditional memory models 43

III. EMPIRICAL DIFFICULTIES FOR ARCHIVAL MODELS 69

3.1 Disputes and impasses in memory science 74
3.2 Revisions to the standard models 96
3.3 Computationalist assumptions, revisited 113

IV. PHILOSOPHICAL DIFFICULTIES FOR ARCHIVAL MODELS 132

4.1 Wittgenstein on memory 132
4.2 A case study of repertory memory 164
4.3 Extirpating Componentiality, Traces, and Hidden Processes 176

V. ALTERNATIVES TO STORAGE 204

5.1 Resources from anti-computational developments in cognitive science 207
5.2 Features in the absence of modularity 215
5.3 Causality and isomorphism in memory 231
5.4 Remembering without storing 250

VI. CONCLUSIONS 256

6.1 Three objections 256
6.2 Implications 265
6.3 Further investigations 272

BIBLIOGRAPHY 276

CHAPTER I

MEMORY AND WITTGENSTEIN

> I saw this man years ago: now I have seen him again, I recognize him, I remember his name. And why does there have to be a cause of this remembering in my nervous system? Why must something or other, whatever it may be, be stored up there in any form? Why must a trace have been left behind? Why should there not be a psychological regularity to which no physiological regularity corresponds? If this upsets our concept of causality then it is high time it was upset.
>
> Ludwig Wittgenstein[1]

> Few problems that inquiring minds have attempted to resolve by the methods of science have resisted understanding as stubbornly as have problems of memory.
>
> Endel Tulving[2]

The study of human memory has encountered deep and sustained conceptual difficulties. Researchers who specialize in the cognitive science of human memory admit this, confessing that we have little, if any, basic understanding of how memory works.[3] Despite this, researchers in other areas of cognitive science, and also in the philosophy of mind, continue to rely—implicitly and explicitly—on models with which memory scientists have become disenchanted. To a remarkable extent, even those scientists and philosophers who explicitly seek to reject accounts of memory based on storage and retrieval fail to avoid the assumptions inherited from these models.

---

1 *Zettel* §610.

2 *Elements of Episodic Memory*, p. 1.

3 Susan Sara: "Virtually nothing is known about the physiological processes underlying the act of remembering" (Sara 2000, p. 76); Jens Brockmeier: "We only have a vague idea about memory as a whole. We cannot even say if there is such a thing as memory (or a memory, or specific memory systems) at all…" (Brockmeier 2010, p. 5); "Yes, we can talk about memory systems and memory processes…but we have little idea how "real" these systems and processes are" (Tulving 2002, p. 323).

In cognitive science and in philosophy, the study of memory appears to be hobbled by basic assumptions that have so far received insufficient investigation.

At the same time, Wittgenstein scholars and other philosophers have been increasingly claiming that Wittgenstein's work offers tools and perspectives that are essential to clearing away conceptual confusions in the study of mind and cognition.[4] Wittgenstein's work, especially in his later years, frequently focused on psychology, explanations of mental phenomena, and the nature of mental processes. Wittgenstein explicitly cautions us about conceptual confusion in psychology and the sciences of mind on several occasions.[5] Despite this, as Peter Hacker notes, "the realization of the importance of applying Wittgenstein's analytic methods to experimental psychology, cognitive science, and cognitive neuroscience dawned late."[6] The value and role of Wittgenstein's thought for the philosophy of mind and cognitive science has been controversial, to say the least, but there is an increasing number of philosophers and scientists who have found value in applying Wittgenstein's lessons to conceptual issues in cognitive science.[7]

This work is at the intersection of these two trajectories. A growing number of memory scientists, and philosophers, have identified the need to investigate and rethink the assumptions that lie at the foundations of our approach to memory. Much work in the philosophy and science

---

[4] See, for example the *Perspicuous Presentations* collection edited by Daniele Moyal-Sharrock (2007), the collected articles in *A Wittgensteinian Perspective on the Use of Conceptual Analysis in Psychology*, edited by Timothy Racine and Kathleen Slaney (2013), or Rom Harre's *Wittgenstein and Psychology: A Practical Guide* (2005).

[5] In *Philosophy of Psychology—A Fragment* (Called Part II of the Investigations until the latest edition), Wittgenstein correlates the "confusion and barrenness of psychology" with the conceptual confusions found in it (§371); other and similar cautionary remarks can be found throughout this work, as well as his *Remarks on the Philosophy of Psychology* (e.g. §290, §292, §1063), and *Last Writings on the Philosophy of Psychology*, (e.g. §686, §777).

[6] Hacker 2013, p. 24. Notably, Hacker is here introducing a recent volume of articles that bring Wittgensteinian lessons to bear on various psychological concepts and phenomena—but none of these articles focus on memory.

[7] Recent examples include Bennett and Hacker 2003, Hutto 2009, Proudfoot 2009, Trigg and Kalish 2011, Boncompagni 2013, the collection of articles in Racine and Slaney 2013, and Sutton 2014.

of mind and cognition relies on unsatisfactory models of memory, and those who study memory directly are well aware of this problem. At the same time, a growing number of philosophers, and scientists, have called for the application of Wittgenstein's ideas and methods to aid in resolving conceptual difficulties in cognitive science. Although Wittgenstein's remarks often explicitly mention memory or many of the concepts relevant to remembering, memory still happens to be one subject that has not been the target of very much sustained application of Wittgenstein's philosophy. I will argue that Wittgenstein has just the lessons that the sciences of memory need, and that the sciences of memory are grappling with just the sort of conceptual confusion to which Wittgensteinian lessons apply.

After spending this first chapter situating the discussion, chapter two will introduce computationalism and memory science, delineating some computationalist assumptions in the science of memory. Chapter three will survey several empirical difficulties that traditional theoretical frameworks in memory science have encountered, illustrating the way that latent computationalist assumptions have kept memory science from overcoming these obstacles. Chapter four will give several philosophical arguments against archival models of memory—several of which stem from Wittgenstein's work, but also drawing on other philosophers concerned with psychological phenomena. These arguments will show how the basic notion of memory as a store or archive is untenable and problematic. Chapter five will build on various existing work on memory that points the way toward a revised framework that is free of these problematic assumptions. In the final chapter, objections, implications, and future directions will be discussed.

In this first chapter, I will survey some of the motivations for such an investigation, laying the groundwork and context for a Wittgensteinian critique of some basic assumptions in the study of memory. In the first part I will survey the role of memory in philosophy at large, showing that investigations into the actual workings of memory are valuable to the philosopher *qua* philosopher. Then I will briefly canvass the scientific study of memory, and especially those developments which have demonstrated that the storage models of memory so ubiquitously and casually employed are insufficient, and which have left various memory scientists calling for attention to basic concepts. This will outline the need for sustained analysis of the concepts of memory.

In the second part of the chapter I will introduce some of the discussion surrounding Wittgenstein's philosophy of mind, and whether Wittgensteinian lessons can be valuably applied to cognitive concepts and phenomena studied in the 21st century. I will show that memory is a good test case for this, and that successfully applying Wittgensteinian ideas to the conceptual difficulties in memory science will clarify and support the ways in which Wittgenstein's views have important consequences for our understanding of mental concepts and phenomena. This will motivate a Wittgensteinian analysis of the concepts of memory.

In the final part of the chapter, I will discuss some recent and parallel developments in cognitive science. As a broadly anti-computational, anti-representational, revisionist turn, rethinking some of our basic conceptual framework in cognitive science, the Wittgensteinian critique and account of memory I will offer sits among and beside other and similar developments in cognitive science from the last few decades. I will survey some of these research programs, movements, and critiques, so better to situate the present project. This will, I hope, help to clarify the aims and place of the present analysis. It is my hope that we will then have good reason, on behalf of philosophers and scientists, to rethink memory, good reason to look to memory as a suitable subject for Wittgensteinian lessons, and an idea of how this reconceptualization fits into the context of contemporary movements in cognitive science.

1.1 Memory in philosophy and cognitive science

From the aviaries of the *Theaetetus*, to the transformation of Hume's impressions into ideas, to the viability of strong A.I. or the extended mind hypothesis, claims about the mechanisms of memory have been central to a broad array of philosophical argumentation. Descartes relied on the mind's ability to turn to the memory traces imprinted on the brain in order to explicate the interaction between mind and body.[8] The disagreement between Locke and Reid as to what constitutes a person is grounded in facts about the operation of human memory. Hume relied on particular characteristics of forgetting and false remembering in order to prove that only the "force and liveliness" of impressions and beliefs distinguishes them from the imagination.[9] Russell, following Hume, declared that the "whole status of images as 'copies' is bound up with the analysis of memory."[10] Dennett claims that the particular models of memory invoked by functionalists are indicative of deep problems with standard functionalism.[11] Chalmers' critique of Searle's Chinese Room is based on the way human memory actually works,[12] but Rupert's critique of Chalmers' (and Clark's) extended mind hypothesis, in turn, is based on the way human memory actually works.[13] Jesse Prinz relies on the particular functions of working memory to provide his account

---

[8] From his letter to Hyperaspistes, August 1641, "The mind, though really distinct from the body, is none the less joined to it, and is affected by traces impressed on it, and is able to impress new traces on its own account. ... The motion of these brain particles leaves behind the traces of on which memory depends" (Descartes 1991, p. 190). See also his letter to Mesland, 2 May 1644, and his letter to Arnauld, 29 July 1648.

[9] This argument is articulated especially clearly in the *Treatise*, in Book 1, Chapter 3, Section 5 (Hume 1738/2007, p. 59-61).

[10] Russell 1921, p. 159.

[11] Dennett 1992, p. 270.

[12] Chalmers 1996, p. 326.

[13] Rupert 2004.

of consciousness.[14] The particularities of memory are bound up with treatments of personal identity, occurrent and non-occurrent belief, artificial intelligence, and the nature of the mental. Memory is taken, along with such illustrious and well-versed examples as perception, language, and rationality, to be one of the central capacities of the mind. Moreover, the particular details of the way that memory works are crucial to the success of various theories that seek to explain, or even merely to invoke, cognitive phenomena.

Despite these roles, direct and sustained philosophical analysis of memory is relatively rare. John Sutton, in his article on memory in the Stanford Encyclopedia of Philosophy, notes that a proper analysis of memory is likely to be important to a range of philosophical issues, yet memory, he writes, "has often been curiously neglected by philosophers."[15] Sven Bernecker, in his recent philosophical analysis of memory, also notes that although memory has long been recognized as a "central component of the mind…memory is a neglected topic in contemporary philosophy."[16] Kourken Michaelian suggests that perhaps the reason that there has been a paucity of work in philosophy on the nature of memory is because "most philosophers have assumed that something close to Martin and Deutscher's (1966) causal theory of memory is right."[17] Michaelian may be right that something *close to* this theory is assumed to be right, but given the relatively low profile of Martin and Deutscher's 1966 article among philosophers, this acceptance

---

[14] Prinz 2012, Ch. 3.

[15] As Sutton recounts, MacNabb writing of Hume said that "the unsatisfactory nature of Hume's account of memory is noticed by nearly all of his commentators. It is a fault however which he shares with nearly all other philosophers" (MacNabb 1962, p. 360).

[16] Bernecker 2009, p.2.

[17] Michaelian 2011b, p. 323.

must be casual and unstudied.[18] As Michaelian points out, the number of book-length treatments of memory by philosophers is staggeringly low. Furthermore, what analysis has been done is in many ways preliminary. After devoting a chapter to the project almost one hundred years ago, Russell confessed, “This analysis of memory is probably extremely faulty, but I do not know how to improve it.”[19] Russell’s sentiment has not been uncommon in the work on memory that has followed. Even a cursory glance at the contemporary literatures in the philosophy of mind and the philosophy of cognitive science reveals broad and rich discussions of, for example, perception, belief, or attention, with little concerted effort toward the analysis of memory.

This is particularly striking with regard to the contemporary literature, as certain assumptions about the way human memory works lie near the center of several lively debates. This has been the case for some time in various traditional threads in the philosophy of mind: the nature of human memory is known to be bound up with a proper analysis of personal identity[20], or with discussions of the nature and role of beliefs.[21] In addition to these, there are several recent developments in the philosophy of mind and cognitive science that leave an important space for an analysis of memory. One example is the flourishing class of current debates concerning the

---

[18] Martin and Deutscher’s article has garnered something in the neighborhood of 125 academic citations in the past half-century. This number compares favorably to that of any other philosophical treatment of memory, but it compares in a *profoundly* disfavorable manner to standard philosophical treatments of other major components of mind and cognition (e.g., perception, rationality, belief, language, consciousness, attention, representation, and imagination, among others).

[19] Russell 1921, p. 187.

[20] The starting point here is Reid’s critique of Locke’s criteria for personal identity, which is grounded in the basic operations of human memory (Reid 1785/2000, *Essay III*, *Chapter 6*), and the discussion continues through 20th and 21st century analyses of personal identity (e.g., Shoemaker 1959, Parfit 1971, Klein and Nichols 2012).

[21] A.J. Ayer recognized that a proper analysis of memory was probably essential to a complete understanding of knowledge (Ayer 1956, chapter 4). Alvin Goldman devoted an essay to the proposal that the epistemology of belief will be helpfully informed by study of the cognitive psychology of memory decades ago (Goldman 1974), and this juncture continues to receive attention from philosophers (see Senor 2014 for a survey of epistemological problems of memory).

nature and role of mental representation. For instance, Tyler Burge claims that "being an individual with representational mental states is constitutively dependent on having a memory that can re-employ some of those mental states."[22] Even setting Burge's particular account aside, it is clear that memory and mental representation are intimately connected issues, and that the characterization of exactly how memory works will inform discussions of representation. Indeed, Burge characterizes the application of these representations as inheriting the "memory files associated with perceptions of particulars," depicting and relying on an account of memory in terms of stored files that can be retrieved.[23]

Burge is not alone in casually utilizing the language and concepts of storage models of memory in his theorizing. The idea that human memory is essentially archival is crucial in much contemporary argumentation. One straightforward example of this is the much discussed argument for the hypothesis of Extended Mind from Andy Clark and David Chalmers: the original argument rests on the suitability of the analogy between a normal human's memories and an Alzheimer's patient's information stored in a notebook.[24] Somewhat more generally, philosophical characterizations of *belief* often assume that such beliefs (or the subjects of such beliefs) can be stored in memory.[25] Kourken Michaelian and John Sutton, drawing on a collection of research, note that theoretical discussions of extended and embedded mind have been paying

---

[22] Burge 2010, p. 63.

[23] *Ibid*., p. 546.

[24] Clark and Chalmers 1998. The fact that Clark and Chalmers rely on an outdated understanding of memory in this argument is illustrated and criticized by Robert Rupert (2005) and, paying even more attention to the details of memory we will encounter below, Paul Loader (2013).

[25] Eric Schwitzgebel writes that the majority of contemporary philosophers of mind accept the bulk of the standard characterization of belief, which relies on the belief being or having a stored token in memory (Schwitzgebel 2014).

insufficient attention to the actual way memory works.[26] Given that memory and its functions turn up in a variety of philosophical issues and arguments, philosophers have good reason to be cautious not to invoke an inadequate account of remembering.

While a survey of the problems in philosophy of mind and cognitive science reveals an implicit need for investigation into the conceptual framework that characterizes memory as storage, a survey of the literature in the science of memory finds the same need explicitly stated. The scientific study of memory spans several disciplines, is pursued at various levels of analysis, and draws data from an array of widely disparate sources. Its hybrid and multiform nature underscores an especially emphasized need for conceptual clarity. To complicate matters further, the phenomena of memory are among those things that are most familiar to us. We cannot but theorize about memory, and our ordinary ways of speaking about it, which are inherited by memory researchers, are laden with haphazard associations and hidden ambiguities.

Many memory researchers are acutely aware of these difficulties. Memory studies as a concerted, interrelated, interdisciplinary collaboration has only come into its own in recent decades, prompting many in the field to explicitly seek conceptual perspicuity. A recent collection of essays edited by memory scientists Henry Roediger, Yadin Dudai, and Susan Fitzpatrick is directed toward just this sort of conceptual clarification. Prefacing these essays, the editors write:

> What, if anything, holds the entire enterprise together? Can we develop a unified science of memory? [The present work] is based on the assumption that the answer to the second question is yes. Further, we assume that at least part of the answer to the first question is concepts. That is, we believe that cross-disciplinary understanding of key concepts represent a critical element in developing a unified science of memory.[27]

[26] Michaelian and Sutton 2013, p.17.

[27] Roediger et al. 2007, p. xi. Scientists are not always so sympathetic to the analysis of basic concepts—this common concern on the part of memory scientists provides a real and somewhat rare opportunity for collaborative efforts among philosophers and scientists.

Whatever the status of the unification of memory sciences, successful communication and fruitful exchange among research programs is certainly to be desired, and cross-disciplinary understanding (and for that matter, intra-disciplinary understanding) of key concepts is a necessary condition for fruitful exchange. About a decade ago, Endel Tulving—one of the foremost living scientists of memory—acknowledged that the "science of memory" as an interdisciplinary, relatively unified, rigorous enterprise was still in its nascent stages. He openly lamented the fact that little systematic conceptual work had been done on memory.[28] The essays contributors to the Roediger, Dudai, and Fitzpatrick volume are not alone: in the last ten years, various memory scientists have responded to this worry by focusing on the concepts being employed, and have reassessed and to some extent reformed their own basic modeling assumptions as a result. Many researchers who work on memory have become well aware that various aspects of the older, computer-inspired storage models are deficient.[29] Disparate research threads, several of which will be explored below, have emphasized the dynamic, reconstructive, fluid, multiform nature of human memory in contrast to the inertness of classical storage models. Nonetheless, the deeply ingrained habits of archival characterizations of memory persist even in research programs that provide results that belie the status of these habits. Memory researchers still labor, in spite of themselves, under the long shadows of computationalist assumptions about memory. Outside of memory research proper, but in other areas of the cognitive sciences, the storage and retrieval

---

28 Tulving adds: "Formal and focused attention paid to concepts is highly likely to result in greater terminological and conceptual clarity of thought and communication in the discipline" (Tulving 2000, p. 34). Curiously, he also there denies "that the absence of any focused emphasis on concepts has been an impediment of any kind" (*Ibid*.).

29 "Few memory researchers, in private, defend the fixed memory trace paradigm. They know that memory is far more dynamic than our models have typically allowed" (Nadel 2007, p. 181).

picture of memory does not even enjoy this conflicted status: it is all too often simply tacitly accepted.[30]

Some philosophers have also noticed this need for conceptual clarification among problematic areas of cognitive science.[31] Timothy Racine and Kathleen Slaney have recently noted that many areas of psychology have encountered sustained difficulties in clarifying what seem like basic questions, and Racine and Slaney have edited a collection of essays that offer a kind of Wittgensteinian conceptual analysis in response to these difficulties.[32] They characterize this analysis thus:

> What is conceptual analysis, and how might it be helpful to social and behavioural scientists? As the volume will suggest, there is not a single or straightforward answer to these questions. However, in our view the most important aspect of conceptual analysis for Wittgenstein is not so much to follow some prescribed set of general methods but rather to develop a certain attitude to philosophical puzzles, including those that arise in psychology. Not an attitude of suspicion, dogmatism or scepticism, but of curiosity, open-mindedness and scrupulous attention to detail. Although it is common, to speak of Wittgenstein's philosophy as containing a set of methods, which it surely does, we would suggest that these methods are a consequence of the Wittgensteinian attitude toward paying careful attention to the concepts that feature in our empirical work and theories.[33]

Racine and Slaney do not mention memory science in particular, however, nor do any of the contributing authors to the volume of essays this passage introduces. This lacuna is all the more salient given that Roediger, Dudai, and Fitzpatick, above, are clamoring for precisely this open-minded attention to the basic concepts at work in memory science. There is good reason for

---

[30] Stephen Kosslyn writes in an influential work on visual imagery of the cognitive components of his model, "each component is also assumed to store information" (Kosslyn 2005, p. 335), and his models explicitly rely on memory as storage.

[31] "To bring clarity to the sciences of the mind and to eradicate conceptual confusions in these sciences is a worthy vocation for philosophy" (Hacker 2013, p. 26)

[32] Racine and Slaney 2013, p.1.

[33] *Ibid*., p. 4.

focused and sustained philosophical analysis of the storage models of memory, given that these are tacitly and importantly accepted in various moves throughout the philosophy of mind and cognitive science, and that memory scientists are explicitly calling for this analysis.

1.2 Wittgenstein and the philosophy of mind and cognitive science

The other strand of motivation comes from Wittgenstein and the application of his work to our understanding of the mind. Much of Wittgenstein's work is particularly and persistently focused on mind and mental phenomena. Indeed, there are few, if any, threads of discourse in Wittgenstein that remain very distant from discussions of the nature of mind for very long. Wittgenstein himself speaks to the pervasive quality of the subjects:

> The treatment of all these phenomena of mental life is not of importance to me because I am keen on completeness. Rather because each one casts light on the correct treatment of all.[34]

Despite this sustained focus on both the contents and methods of investigations into mentality, there is lingering controversy as to whether there are valuable lessons from Wittgenstein for contemporary sciences of mind, and if there are, just what forms these might take. He is also often cited as one of the primary figures in 20th century philosophy of mind and language, and yet there is little agreement as to just what his contributions to our understanding of the mind consist in. Commentators acknowledge the breadth and richness of Wittgenstein's writings on the mind, but also worry that the lessons of these have gone unheeded. Anthony Kenny writes in his introduction to Wittgenstein of his disillusionment in this regard:

> I had imagined that once his philosophical ideas had been absorbed, thinkers in various disciplines would begin to apply them, with beneficial effect, to work in their own field. ... Some of Wittgenstein's insights into philosophy of language and philosophy of mind, I had believed, constituted irreversible advances in the subject. But now they were obscured or

[34] *Zettel*, §465.

> forgotten. Metaphysical weeds that his probing should have rooted up once and for all returned in ever more abundant strength.[35]

While Kenny's is a particularly pointed example, the sentiment is not uncommon among those philosophers influenced by Wittgenstein.[36] Wittgenstein's work has stirred up intricate controversies and literatures in certain areas of philosophy, yet there has been a lack of application of his ideas to those external research programs that, each in their own way, employ some of assumptions that are the very targets of Wittgenstein's critiques.

Attempts to rectify this have been recently developing, especially in various research areas within and orthogonal to cognitive science. One notable example is Hacker and Bennett's (2003) critique of cognitive neuroscience, in which the authors attempt to show not only that there are underlying assumptions in the cognitive neurosciences that run afoul of various Wittgensteinian insights. Bennett and Hacker also isolate particular work and ways in which these assumptions are manifest, and to offer alternative conceptual frameworks. In response, John Searle also characterizes the "Wittgensteinian vision" of mind, in order to point out just where Bennett and Hacker have gone wrong.[37] As it happens, these characterizations seem to have led to impasse: Bennett and Hacker take the Wittgensteinian vision to withhold from neuroscience the "cognitive", while their critics take the fact of the cognitive in neuroscience to disconfirm the "Wittgensteinian vision." Hacker and Bennett's book is not alone: there are several projects currently afoot that explicitly seek to apply Wittgenstein's ideas to work in cognitive science, both in general and with

---

[35] Kenny 1973/2006, p. xii.

[36] Sluga 1996, p. 321; Putnam 1999, p. 45; Hacker 2013, p. 24.

[37] Searle 2007, p. 101.

particular scientific enterprises in view.[38] Commentators have point out, rightly in my view, that studies in cognitive science, as well as studies of Wittgenstein, stand to gain from further cross-pollination.[39]

In addition, there are also several lines of critique within the philosophy of cognitive science, some of them longstanding, that draw more broadly on Wittgenstein's ideas.[40] These concrete applications of Wittgenstein's work to present work in cognitive science are important projects, not only for the beneficial reconceptions their purveyors hope to provide to those doing and thinking about the cognitive science in question, but also for the elucidation of Wittgenstein's contributions to our understanding of mind and mentality. Commentators with broadly Wittgensteinian inclinations have insisted for decades, like Kenny, that Wittgenstein's work offers profound and irreversible advances in the ways we characterize and understand mental phenomena. Various critics have resisted these claims, and Wittgenstein's status in the philosophy of mind and cognitive science remains contested. One important component in coming to a clearer understanding of the implications of Wittgenstein's philosophy is just this application to particular current work in the sciences of mind, and the way that particular cognitive capacities are presently understood.

Memory is a nearly ideal candidate for this. Wittgenstein's work contains numerous remarks on memory, some of which directly engage with what has become current cognitive

[38] Meyers and Waller 2008; Hutto 2009; Susswein and Racine 2009; Trigg and Kalish 2011.

[39] Anna Boncompagni writes, "What I am interest in suggesting, however, is that a Wittgensteinian framework for cognitive science is a perspective worth working on" (Boncompagni 2013, p. 28). *Cf.* Sutton 2014.

[40] For example, enactive perception (e.g., Noë 2004, Boncompagni 2013, Preston and Schroeder 2013) or Dreyfus' critiques of artificial intelligence (Dreyfus 1972/1992, Dreyfus 2007)—each of these will be discussed below.

science.[41] The processes and mechanisms of remembering appear as subjects of direct philosophical concern especially throughout those later writings of Wittgenstein's that are published as the *Remarks on the Philosophy of Psychology*, *Zettel*, *Philosophy of Psychology—A Fragment* (once known as "Part II" of the *Investigations*), and *Last Writings on the Philosophy of Psychology*. Moreover, the arguments and concerns in the *Philosophical Investigations* are never very distant from the phenomena of human memory: the proper characterizations of learning, recognition, or recalling images play major roles throughout the work. If one is looking to apply Wittgenstein's insights about the mind to the study of particular cognitive capacities then memory is a natural choice. This claim is not new: researchers have noted that Wittgenstein's work is directly applicable to the study and characterization of memory. Norman Malcolm wrote a book on the subject decades ago and various authors have approached it intermittently since.[42] These are still relatively few in number, however, and treatments of memory are conspicuously absent or brief in recent collections of work that applies Wittgensteinian ideas to philosophical psychology and cognitive science.[43]

The application of Wittgenstein's thought to the cognitive science of memory deserves more attention.[44] There is a particularly Wittgensteinian merit in applying these philosophical insights directly to work on *particular* cognitive capacities rather than trying to construct a "Wittgensteinian" theory of mind without looking at how psychological concepts are used. As a

---

[41] These will be enumerated and elaborated below. Notable remarks can be found, e.g., in the *Blue and Brown Books*, pp. 85-88, 182-185; *Philosophical Investigations* §§56, 265, 305-306, 343; *Zettel* §§25, 136, 202, 610-612.

[42] Malcolm 1977; Stern 1991; ter Hark 1995; Frongia 1995; Section 5.2 of Bennett and Hacker 2003; Hamilton 2007; Moyal-Sharrock 2009.

[43] Moyal-Sharrock 2007 and Racine and Slaney 2013 are both collections of articles that approach many aspects of cognition using Wittgensteinian strategies and arguments; neither includes any sustained treatment of memory.

[44] *Cf.* Boncompagni 2013, Sutton 2014.

particular cognitive capacity, memory is a good candidate for this: Wittgenstein was directly and indirectly concerned with the processes of remembering, and several debates in modern memory science mirror Wittgenstein's concerns. A Wittgensteinian critique and reconception of standard, archival, memory models will aid both those researchers who seek to understand the mechanisms of memory and those researchers who aim to characterize Wittgenstein's contributions to our understanding of mind and cognition.

1.3 New directions in cognitive science

This project criticizes certain basic assumptions tacitly accepted in various research programs within and near to cognitive science on both philosophical and empirical grounds, and also attempts to offer an alternative conceptual framework that will allow us to adequately characterize the phenomena in question. As such, it exists in the company of a broad array of projects in the philosophy of cognitive science, especially a collection of relatively recent developments.[45] It will help to better understand the present project to situate it with respect to these philosophical movements within and surrounding the last few decades of cognitive science.

Some of these are well-defined research programs with a long history. Hubert Dreyfus has been successfully offering Heideggerian/Wittgensteinian critiques of, and alternatives to, classically computationalist A.I. for several decades.[46] Dreyfus' critique is multifaceted, but at least one of the continuing themes in this critique is directly applicable to memory. Dreyfus critique often relies on the apparent nature of *expertise*. When humans are expert at something,

---

[45] It is out of consideration for these movements that William Ramsey writes that the cognitive revolution is now "moving backwards" (Ramsey 2007, p. 223).

[46] The starting point for this critique is Dreyfus' What Computers Can't Do (1972). Many of the same threads figure prominently throughout Dreyfus' work (e.g., Dreyfus 1981, Dreyfus 2007), and they have been picked up in other work which follows in this tradition (e.g. Hutchins 1995, Ch. 9; Bickhard and Terveen 1996; Dennett 2006).

the phenomenology of their expert agency seems completely free of inference, algorithmic rule-following, or effortful recall. Rule-following, deliberate recollection, and inference are, as Dreyfus points out, marks of the novice. Paradigm cases of intelligent action are those in which the agent simply *responds*, without following rules or executing complicated searches and algorithms.

Although Dreyfus' primary target is a class of traditional strategies in artificial intelligence research, it is not difficult to see how this point about expertise has implications for memory. In fact, A.J. Ayer had already made this point about remembering in particular, in the chapter on memory from *The Problem of Knowledge*.

> It can indeed happen…that people are assisted by actually recalling some previous occasion on which they did the thing in question, or saw it done, but it is by no means necessary that they should be. On the contrary, the better they remember, the less likely it is that they will have any such events in mind: it is only when one is in difficulties that one tries as it were to use one's recollections as a manual. To have learnt a thing properly is to be able to dispense with them.[47]

Ayer goes on in this chapter, in what might even be called a Wittgensteinian fashion, to use these considerations to show that there need be no inner process that accompanies remembering. In so doing, Ayer is making the point later associated with Dreyfus' critique not just *by means of* the nature of remembering, but *about* the nature of remembering. I will return to Ayer's argument—as well as to Dreyfus—below. Whereas Dreyfus invokes the nature of learning, memory, and expertise in order to caution researchers against an excess of computationalism in characterizing (and attempting to engineer) intelligent action, the same points about learning and memory can be, and shall be, invoked in order to caution researchers against an excess of computationalism in characterizing memory itself.

---

[47] Ayer 1956, p. 135.

There are other relevant philosophical movements within cognitive science that are developing research programs with a particular focus. Alva Noë has been, in recent years, offering philosophical and empirical critiques of our basic assumptions about perception and purveying an alternative conceptualization of what it is to perceive, drawing on Merleau-Ponty, Wittgenstein, and various contemporary research in the cognitive science of perception.[48] Noë draws on arguments that follow Merleau-Ponty and Wittgenstein in order to show that it is not the case that perception must be a representational, inferential process that ranges over independent and static component parts. He then draws on a variety of recent data from the science of perception (especially vision) to show that what we know about how vision actually works fits poorly with the computation-inspired traditional view of perception. Taking these considerations together, Noë argues for a major revision to the way we think about perception. We have been misled, in part due to the influence of the computer metaphor, into characterizing vision and perception as all too computational.

Noë's enactivist project in the science and philosophy of perception runs parallel to what this project attempts to achieve for the science and philosophy of memory. The arguments Noë takes from Wittgenstein (and, to some extent, those he takes from Merleau-Ponty), to show the objects of our perception need not be inner representations that in turn reflect the world, have analogs in a study of memory. The way that empirical results from the science of vision are shown to be uncomfortably juxtaposed with the precepts inherited from computationalism will have close parallels with the poor fit that will be revealed between empirical results from the science of memory and the computationalist assumptions that still lie in its foundations.

---

[48] Noë's work is a continuation of earlier work by and with Evan Thompson and Francisco Varela (e.g., Varela et al. 1991, Noë et al. 2000), but Noë's later distillation of Thompson and Varela's ideas has become a rallying point for enactivism in perception. A good starting point is Noë's *Action in Perception* (2004).

Other kindred movements are more scattered and sprawling: recent movements like *embodied* cognition or *situated* cognition encompass assorted projects and programs (even including, roughly, the above examples) that emphasize the non-neural components of cognition. Various research programs in cognitive science, and movements in the philosophy of cognitive science and the philosophy of mind, have been exploring and purveying interrelated theses and hypotheses concerning embodiment and situatedness in cognition. Lawrence Barsalou categorizes all of these disparate programs under the heading of *grounded* cognition, emphasizing the different modes and physical systems in which cognitive processes are found to inhere.[49] Julian Kiverstein and Andy Clark divide these into embodied, embedded, and extended approaches to cognition, but admit that the boundaries between these may be vanishing.[50] Richard Menary includes the enactive approach to cognition along with embodied, embedded, and extended approaches under the "4E" umbrella—though he explicitly excludes distributed and situated approaches.[51] Others prefer the name 'distributed cognition' for some of these research programs, emphasizing the parity among neural and environmental elements of cognitive systems.[52]

What ties many of these programs together is just that they offer criticisms of one or more of the fundamental theses of the classical computationalism that is has been the dominant paradigm for the philosophy and sciences of mind throughout much of the twentieth century. Anna Boncompagni writes, "A unifying factor of these accounts is their common opposition to traditional cognitivism, the latter usually being described as the study of cognition seen as a mental

---

49 Barsalou 2008.

50 Kiverstein and Clark 2009.

51 Menary 2010, p. 459.

52 Hutchins 1995; Sutton et al. 2010.

computation based on inner representations."[53] Some of these approaches to cognition invoke alternative cognitive architectures, particularly those associated with connectionism and dynamical systems theory, which in turn also count among the examples of recently developing movements in the philosophy of cognitive science. Connectionism has a long and colorful history in cognitive science,[54] while the dynamical systems approach to cognition has been more straightforwardly gaining ground for the last two decades. Recently there have been a few bold attempts to synthesize insights from both of these traditions with theoretical frameworks from cognitive psychology.[55]

Several of these projects and research programs do touch on memory and its place in cognition, even if none have focused on memory in an explicit or sustained fashion, and memory science has developed somewhat independently in the last few decades.[56] An account of memory that avoids the pitfalls of storage and retrieval will be compatible with, and helpful to, various projects in embodied, enactive, or situated cognition. The positive account of memory I seek to construct and advocate later in the project will be drawing from several of these projects—

---

[53] Boncompagni 2013, p. 28. *Cf.* Menary 2010, p. 460. It is worth noting that some of these research programs explicitly embrace versions of computationalism—Hutchins, for example, is a prominent advocate of situated, distributed cognition but characterizes cognition as "computation realized through the creation, transformation, and propagation of representational states"—these computations are simply distributed socially and environmentally (Hutchins 1995, p. 50).

[54] Famously, neural networks were subsequently ignored for several decades after their mid-twentieth century discovery. From the late 1980's to the mid 1990's they enjoyed a meteoric renaissance, but philosophical debates about the properties of neural networks seem to have stalled out in the 1990's despite continued technical work on networks (Garson 2012). Nominal endorsement of a mild connectionism is now widespread throughout cognitive science, though many of the crucial concerns about the implications of neural networks for cognitive science remain unresolved.

[55] Dynamical systems approaches have been closely tied to conceptual frameworks in developmental psychology since their inception (Thelen and Smith 1996). For examples of successful—albeit preliminary—integrations of problems from cognitive psychology, analyses of neural networks, and dynamical approaches, see the articles collected in Spencer et al. 2009b.

[56] Michaelian and Sutton 2013, p.7.

particularly from existing work on memory in neural networks and dynamic systems theory.[57] There are many recently developed tools and resources for doing cognitive science outside traditional boundaries. There are several new methodologies and conceptual frameworks for understanding cognition without relying on the computer metaphor as thoroughly as traditional approaches have done. Several particular subjects of research—perception, artificial intelligence, human development—have already become loci of revisionist movements in cognitive science. Given that many scientists working within the study of human memory are explicitly calling for conceptual investigations and dramatic revisions to standard frameworks, a critique of the standard approaches to memory is overdue.

[57] Two examples of useful constructive (as opposed to that which is purely critical) work on memory and alternative architectures are McClelland 2000 (connectionism) and Spencer et al. 2009a (dynamic systems). A preliminary but promising attempt to model data from the cognitive neuroscience of memory using connectionist and dynamic techniques can be found in Kryukov 2008.

CHAPTER II

COMPUTATIONALISM AND MEMORY

> Our views of the operation of memory are fuelled by the procedures and techniques we have invented for the preservation and reproduction of information. … The history of memory is a little like a tour of the depositories of a technology museum.
>
> Douwe Draaisma[58]

> It is one of the great intellectual achievements of the 20th century that we are able to do so much with such a simple apparatus, but equally it is one of the great intellectual mistakes of the latter part of the 20th century to suppose that that is what is going on in our minds.
>
> John Searle[59]

Accounts that characterize memory as an archive have been predominant throughout the history of memory research and inquiry.[60] The digital computer is only the latest in a long chain of metaphors that cast memory in terms of storage and retrieval.[61] All of the present argumentation and work would probably still be relevant whether or not cognitive science had ever drawn so much from computationalism, and even today most of the following critiques could be made without explicit mention of computers or the computational theory of mind. Many of the underlying assumptions that I will be attempting to undermine existed long before computers did, and long before minds or memories were conceived of as computational. Still, computationalism has neatly distilled these assumptions, and understanding these models as computationalism-inspired will help focus the investigation. Furthermore, there is ongoing debate concerning the

---

58 *Metaphors of Memory*, p. 3.

59 *The Mind and Education*, p. 7.

60 Brockmeier 2010, p. 7.

61 See Draaisma 2000 or Roediger 2000 for thorough surveys of these metaphors.

merits and demerits of computationalist assumptions in the philosophy and sciences of mind, and the following characterization of these problematic assumptions lying in the foundations of memory research may contribute to an understanding of computationalism as a whole. I will therefore briefly characterize computationalism in its modern guise, as well as some of the influential anti-computational critiques in recent philosophy of cognitive science. Some of the critics of computationalism have isolated particular and problematic assumptions that are still in play, even among researchers who do not explicitly count themselves as computationalist. After this sketch of computationalism, I will give a basic description of mainstream accounts of memory and the conceptual frameworks these exist within. Lastly I will isolate three computationalist assumptions that remain foundational to much work in the study of memory.

2.1 Computationalism and its legacy

To the untrained eye, it can appear as though the power of the computer metaphor waxed and waned during the twentieth century. Computationalism—the view that we should understand cognition as computation—was once the only game in town.[62] In recent decades, the sciences and philosophy of cognition have begun to pursue varied and vibrant alternative research programs. Overt advocacy of computationalism in psychology or philosophy, has become comparatively scarce.[63] The aim to achieve human-level artificial intelligence solely by means of classical computation has been all but abandoned.[64] Computational models of vision and perception have

[62] In 1975, Fodor could still reasonably claim that the computational theory of mind was the only "remotely plausible" strategy on offer, that it was the "only game in town" (Fodor 1975). By the mid-eighties Fodor and other champions of a computational view of mind had to rely on other arguments. A parallel narrative runs through explicit avowals of computationalism is psychology (Roediger 1996, p. 82).

[63] Randall 2007, p. 614.

[64] See Gryz 2013 for a discussion of this abandonment as related to the frame problem.

been replaced or dramatically refined.[65] Various recent movements in the philosophy of mind have emphasized the non-computational, especially organic aspects of our cognition.[66] Compared to the wide currency that explicitly computational models and theories enjoyed two or three decades ago, the present-day potency of the computer metaphor seems radically diminished. With the exception of a few holdouts who persist in claiming that computationalism is still worth talking about,[67] popular sentiment among cognitive scientists and philosophers might be characterized simply by noting some of the titles of the last few popular articles to mention the subject in the 20th century: "The Failures of Computationalism", "What Could Cognition Be, If Not Computation", "Beyond Computationalism", "Computationalism is Dead. What now?"[68] John Searle called computationalism "one of the great intellectual mistakes of the latter part of the 20th century."[69] It can, at first glance, seem like philosophy of mind and cognitive science turned over a new, postcomputationalist, leaf with the advent of the 21st century.

Late-twentieth century developments have certainly complicated whatever "mainstream" account may have once been available of the relationship between computation and cognition. Nonetheless, computationalism is not dead. Explicit denials of computationalism may have become more common than explicit advocacy, and a series of arguments against computationalism are sometimes taken to have been decisive, "Yet," as Gualtiero Piccinini notes, "computationalism

---

[65] O'Regan & Noë 2001; Wilson 2010; Orlandi 2011.

[66] Many of these were surveyed in the previous chapter, but the dynamic systems and radical embodied cognition movements are particularly suitable examples.

[67] Notable among these are Gualtiero Piccinini and Jack Copeland, each of whom continues to advocate for a more thorough discussion of the ideas surrounding computationalism in the philosophy mind and cognitive science (Piccinini 2009, Piccinini 2010a; Copeland 1996, Copeland and Proudfoot 2010).

[68] Searle 1993; Van Gelder 1995; Giunti 1996; Bringsjord 1998.

[69] Searle 1998, p. 7.

has become and remains the dominant theory of cognition in psychology, neuroscience, and philosophy of mind."[70] One of the reasons for this apparent double life is that computationalism, as a broad thesis about the mind, has been vague and mutable enough that it has been able to fulfill both of these roles. As Piccinini points out, without due caution computationalism can easily be characterized (and has been characterized) in too trivially weak or too stringent (and clearly false) ways.[71] It has been surprisingly difficult to rigorously spell out a version of computationalism that is neither obviously true nor obviously false, while retaining the spirit of the thesis that was so influential in the twentieth century.[72] This is an important and ongoing project, but using 'computationalism' to refer to a collection of theses that emphasize similarities between computers and minds will presently suffice. I will be focusing on another reason for the computer metaphor's ability to simultaneously play the roles of villain and hero in twenty-first century philosophy and cognitive science: the fact that reliance on computationalist assumptions about the mind has become implicit and unacknowledged in many research areas.[73]

Drawing out implicit and computationalist assumptions about the mind that have been dogging philosophical or scientific research is not, in itself, a new endeavor. Hubert Dreyfus was doing something very close to this decades ago, in the heyday of computationalism. Dreyfus offered a sustained criticism of research programs in artificial intelligence and their reliance on

---

[70] Piccinini 2010b.

[71] Piccinini 2009. Noë agrees (Noë 2009, p. 160).

[72] Horst 2011; Piccinini 2010a.

[73] Hilary Putnam wholeheartedly (and somewhat polemically) agrees: "It is a profound mistake to equate serious science with the Cartesianism cum materialism that has for three centuries tried to wrap itself in the mantle of science. Today that attempt often takes the form of empty talk about "the conceptual structure of the mind"—talk that simply takes for granted the idea that thinking is syntactic manipulation of symbols. Nothing in the successes of serious psychology or linguistics endows that view with content. Instead, such talk frequently lowers the level of philosophical discussion to that of popular "scientific" journalism" (Putnam 1999, p. 48).

unstated (and sometimes stated), computational assumptions about the mind, arguing that the failures of artificial intelligence are indicative of essentially non-computational aspects of human cognition.[74] Dreyfus uses problems in artificial intelligence research to highlight several assumptions: the "computer-influenced assumption that experience can be analyzed into isolable, atomic, alternative choices," and the "idea that feelings, memories, and images must be the conscious tip of an unconscious frame-like data structure," among others.[75] The reception Dreyfus' arguments have received in both the philosophical and scientific communities has been mixed, but it is certainly the case that research in artificial intelligence has been deeply influenced by these arguments. Dreyfus is not only addressing the researcher in artificial intelligence, however, his argument challenges basic and widely held assumptions in the philosophy of mind as well. Steven Horst points this out in his article on the computational theory of mind in the Stanford Encyclopedia of Philosophy:

> However, Dreyfus's argument is not purely inductive; it also contains a more principled claim about the nature of expert performance and the unsuitability of rule-based techniques to duplicating that performance. This argument is a complicated one, and has not received decisive support or refutation among philosophers of mind.

Part of the reason that the argument has not yet received decisive support or refutation is, once more, the fact that self-avowed computationalists have become rarer in the ensuing decades. This is not to claim that Dreyfus' arguments no longer deserve attention: the assumptions targeted by these arguments are widely influential, inside and outside of explicitly computationalist contexts. Indeed, the assumption that conscious remembering is the tip of an unconscious data structure, as we will see, is still an assumption worth considering.

---

[74] Dreyfus 1972/1992.

[75] *Ibid*., p. 165; p. 181.

A more recent series of arguments that draw on empirical work to challenge computationalist assumptions about the mind can be found in work on perception—notably in the recent work of Alva Noë[76], but also among researchers who more explicitly embrace a Wittgensteinian approach like John Preston and Severin Schroeder.[77] Noë has persistently challenged the view that perceptual experience is a product of detailed representations encoded in neural correlates. Although few of the targets of Noë's critiques identify their theories as explicitly computational, at least two of the basic assumptions that, according to Noë, undergird these theories are recognizable inheritors of computationalist views. First, there is the assumption—the increasingly controversial assumption, in recent years—that perception *must* depend on inner representations of the perceived world. Perceptual input is incomplete, distorted, discontinuous, and coarse-grained, goes the common line of thought that Noë challenges, and so the only mechanism by which the apparent unity and continuity of perceptual experience can be achieved is the construction of an inner model. We experience only our input, and so if our experience does not seem to include qualities inherent in this input, we must experience *modulated* input. There must be an inner representation between us and our input, which in turn stands between us and the world. One of the aims of work like that of Noë is simply to show that this way of characterizing our perceptual experience is stilted and artificial. To assume that perceptual experience is, at bottom, a matter of discrete input, isomorphic representation, and internal state transformations is just to assume that cognition is fundamentally information processing, and very computer-like. We do not experience our input. We experience the world. At least, as Noë often insists, the notion that we are separated from the world by input and representation is no foregone conclusion. The

[76] Noë 2004 and Noë 2009 are summary presentations of these critiques; Noë develops on earlier work by Francisco Varela and colleagues (Varela at al. 1991), among others.

[77] Preston and Schroeder 2013.

assumption that our cognition must be a product of input, transformation, and representation is just another latent computationalist assumption.

In addition to the assumption that we, or our brains, must represent that which we perceive, Noë and others in his critical tradition also highlight and question the assumption that internal state transformations range over neural correlates that match the experienced perceptual content. That is, if our perceptual experience is partly constituted by the world and our active part in it, then we must not assume that there will be a mapping of this experience in the brain. As Noë concludes, no neural correlate alone will be enough to produce a given percept:

> We have challenged the minimal neural substrate thesis by emphasizing the active and attentional character of perceptual experience. If the content of perceptual experience depends crucially on the environment, as well as on skillful motor capacities and capacities for directed attention on the part of the perceiver as a situated agent in the environment, then it cannot be assumed without argument (as the NCC programme does) that there is any such thing as a minimal neural substrate sufficient to produce conscious experience. Rather, the substrates of consciousness — in particular of visual perceptual consciousness— seem to cut across the brain–body–world divisions.[78]

Noë is merely the latest in a long line of theorists who have doubted whether it is necessarily the case that mental and experiential phenomena can be mapped isomorphically to brain events. The assumption that cognition proceeds through transformations across neural correlates of cognitive-level states and events is often implicitly or explicitly maintained in philosophy, psychology, and cognitive neuroscience. This is just part of what neuroscientists mean by 'computation'—even though, as Piccinini points out, "it's not clear that what most neuroscientists mean by 'computation' is the same as what most psychologists and computer scientists mean by it."[79] Neuroscientific explanations invoke computation, and cognitive explanations invoke computation, but even researchers who express optimism about both of these concede that any parallel evolution and

---

[78] Noë 2005, p. 25.

[79] Piccinini 2009, p. 519.

convergence among these levels of explanation is still a long way off. John Bickle and colleagues express just such guarded optimism:

> A crucial aspect of the total picture gets neglected: the relationship between the levels, the 'glue' that binds knowledge of neuron activity to subcellular and molecular mechanisms, network activity patterns to the activity of and connectivity between single neurons, and behavior to network activity. This problem is especially glaring when we focus on the relationship between 'cognitivist' psychological theories, postulating information-bearing representations and processes operating over their contents, and the activity patterns in networks of neurons. Co-evolution between explanatory levels still seems more like a distant dream rather than an operative methodology.[80]

Indeed, research in memory is one domain that offers some tantalizing glimpses of the apparent integration of levels, and much of the framework for this integration is computationalist. Past experience "encodes" memories into a neural trace, according to the standard models in memory science. The encoded neural trace persists in the brain until it is decoded by a cognitive retrieval process, which is undergirded or implemented by the corresponding neural process. These processes take place in particular brain regions, and some of them can be characterized in both neural and cognitive terms.[81] Glancing over some of the literature in the cognitive neuroscience of memory, one might be left with the impression that the integration of the cognitive and the neural is all but complete, and that whatever computationalist tenets have informed these models have contributed to their successes.

It is my contention that this is exactly wrong. A study of the concepts at work in memory research will reveal deep and persistent confusions, and it will also reveal that these confusions result from the latent computationalism that lies at the foundations of standard approaches to memory. The unresolved difficulties in the science and philosophy of memory are not details, soon to be reconciled with our models. Rather, these unresolved difficulties are direct

[80] Bickle et al., 2012.

[81] Indeed, some of the neural processes can even be redescribed, to a large extent, in molecular terms.

consequences of some of the basic assumptions inherent to much theorizing about memory. To complicate matters, a study of these assumptions and their treatment in memory science will demonstrate that computationalism in the study of memory has often been half-rejected. That is, philosophers and scientists who study memory are not unaware of the problematic consequences some computationalist bases have brought to particular models. The response to this, I will argue, has been half-hearted. Some computationalist theses about memory have undergone revisions that are merely cosmetic; others have been revised in part but maintained in part. The result of this half-hearted revision is that many models of memory are at odds with themselves, containing internal tensions that manifest as impasse and persistent conceptual confusion. It will be my contention that only a more complete rejection of computationalism in the science and philosophy of memory will allow us to overcome these obstacles.

Understanding computationalism can help to understand these impasses; the computer metaphor makes clear some of the assumptions about memory which have been adopted by researchers across a wide array of disciplines and inquiries.[82] Computationalism as a thesis or as a family of theses is notoriously difficult to define, but for the present purposes we need only a working definition—it is the assumptions derivative from these theses that are ultimately our target. We might then define computationalism as the view that human cognition is robustly computer-like, transforming input to output by performing algorithmic operations over discrete, representational structures.[83] I will outline certain assumptions about the nature of memory that

[82]Computationalism is sometimes characterized in terms of "information processing". As many commentators have pointed out, it is difficult to articulate this notion so that it is both rigorous and interesting. See Piccinini 2009.

[83]This tracks Piccinini's "classical computationalism" pretty closely (Piccinini 2009), and on a strong reading of 'robustly' it also tracks Horst's "computational theory of mind" (Horst 2011). Many characterizations and supporters of this and related views descend from the "physical symbol systems" of Newell and Simon (1976). Compare also Lawrence Shapiro's recent characterization: "An examination of some paradigm work in standard

stem from this picture of the mind, and if certain revisions of computationalism are able to avoid these assumptions, so much the better for these modified forms of computationalism. The target here is the hidden and problematic assumptions in the concepts brought to bear on memory, whatever view or class of views might be said to give rise to them.

Of course, the twentieth-century reliance on the computer metaphor produced and engrained many underlying assumptions across a collection of domains in cognitive science, psychology, and philosophy of mind. The sciences of memory are no exception: computationalist, information processing approaches to the study of memory were pioneered in the 1950's and then largely guided research for the rest of the 20th century.[84] Taking the digital computer as our guide, characterizing memory in terms of storage and retrieval is quite natural. For a computer, any input that will be efficacious in any future transactions must be *stored* in memory.[85] The stored item may be encoded, and can be retrieved in various ways. Tellingly, computer scientists who work on better ways to store and retrieve encoded information in computer memory sometimes call these *knowledge representation* systems. The computer is able to distinguish unfamiliar inputs from familiar ones, to apply old information to new problems, and to reproduce information as this information was input in the past. Additionally, various computer architectures include different types of memory, some of which are suited to long-term storage of versatile information and some of which are suited to keeping track of the threads of input and activity going on in the present, so to speak. It is easy to see how the structure behind these capacities might be appropriated without

---

cognitive science reveals commitments to a computational theory of mind, according to which mental processes proceed algorithmically, operating on symbolic representations" (Shapiro 2011, 27).

84 Roediger 1996, p. 82.

85 In fact, I can still recall when 'computer "memory"' was written with scare quotes.

hesitation by those looking to model the operations of the human mind. Indeed, it has been so appropriated.[86]

There are a number of computationalist assumptions lying near the core of much theorizing about memory even up to the present day. These assumptions about the way that human memory must work impede clarity and progress in the study of memory. By identifying and characterizing these tacit theses in memory research, we may be able to understand what memory models in the absence of computationalism would look like, and how these might overcome many of the problems generated by our present models.

2.2 Modeling memory

Before isolating and articulating some of the basic assumptions common to most accounts of memory, it is necessary to sketch these accounts. What follows is a brief first-run description of some characteristic features of mainstream models of memory. As is the case with all brief and characteristic accounts, it will be simplistic, and will probably not include any single claim that is unanimously agreed upon in the field of memory science. Nonetheless, scientists who study memory—at the neurobiological level, the behavioral level, or anywhere in between—operate within a common framework that has at least some reasonably identifiable features. I will focus on three areas of these models, each of which will merit our attention later in this investigation. The first is the division of memory into different types of remembering. In memory science, this division is usually spoken of in terms of dividing the process into different *memory systems*. The second is the basic, traditional model of the stages of the remembering process. In each instance

[86] Daniel Schacter writes, "Clearly, the ability merely to store and retrieve information is not a unique feature of the human mind, or even of living organisms. Every time we type on our personal computer, we interact with a formidable memory system" (Schacter 2008, p. 34).

of remembering, there is some past event, some present remembrance, and some connection between past and present; each of these stages is represented in the tripartite models that dominate accounts of memory. The third area that I will focus on is the memory trace, or engram. Accounts of memory that talk of *storage* have usually assumed that the stored, encoded thing can be found in the brain, or at least that a neural correlate of this stored thing is encoded in the brain. Although explicit accounts of this memory trace vary widely, there are few researchers who study or characterize memory in any way without at least implicitly committing themselves to such a trace.

2.2.1 *Ways of remembering: memory systems*

The tradition of explaining memory by way of dividing it into processes is a long and variegated one. Descartes, perhaps fittingly, posited one memory system for remembering material things, and another memory system for remembering intellectual things. The former involves traces imprinted on the brain and then later read by the mind, and the latter involved traces imprinted directly on the mind, albeit in a somewhat mysterious fashion.[87] Aristotle's work uses two different concepts of remembering—memnesthai and mimneskein—but it remains controversial how these cut across modern concepts of remembering, or even whether he meant any distinction at all by utilizing these two terms.[88] The philosophers' divisions of memory that modern memory science finds useful start with Bergson and Russell. Bergson divided "habit memory" (the memory for skills and learned procedures) from "pure memory" (the memory for

---

[87] From the Letter to Mesland, May 2nd, 1644: "As for memory, I think that the memory of material things depends on the traces which remain in the brain after an image has been imprinted on it; and that memory of intellectual things depends on some other traces which remain in the mind itself. But the latter are of a wholly different kind from the former, and I cannot explain them by any illustration drawn from corporeal things without a great deal of qualification" (Descartes 1991, p. 233).

[88] Sorabji 1972.

past events, episodes or facts).[89] Russell followed Bergson and further elaborated the differences between remembering how to do things and other types of remembering.[90]

Remembering how to do something is, in many ways, quite different from remembering that something is the case. We may remember how to play the mandolin, or we may remember that Trieste is in Italy, but even the merest reflection on the way we talk about, use, and experience these abilities makes it clear that the word 'remember' is importantly bifurcated between these instances. These simple and apparent differences provide one motive for the division of memory into memory systems.[91] The classification of distinct memory abilities is still far from a settled matter, but one distinction that continues to be fruitfully invoked—and refined—is that between declarative and nondeclarative memory. That is, if I remember that Trieste is in Italy, or recount the oral defense of my comprehensive examinations, or recall that the Berlin Wall fell in 1989, the expression of my memory is primarily verbal.[92] On the other hand, if I can remember how to play the mandolin, or how to skip a stone, or how to distinguish beeches from elms, the expression of my memory is primarily non-verbal—in fact, non-declarative memory is often particularly ill-suited and resistant to verbal expression. Remembering facts, propositions, or events requires declarative memory, whereas remembering skills, competences, or procedures requires nondeclarative memory. While a rigorous characterization of this distinction has not been without

[89] Bergson 1896/2004.

[90] Russell 1921, Ch. 9.

[91] The other primary motive for this has been the asymmetry present in the competences of densely amnesic patients--a point to which we shall return. For a review of memory systems in general, see Schacter and Tulving 1994. For a recent philosophical study of the implications of memory classification, see Michaelian 2011.

[92] The precise nature of the relationship between speech and declarative memory continues to provoke deep controversy--for present purposes I seek to remain as neutral as possible on this. At the very least, declarative memory, unlike nondeclarative memory, is usually expressed in words if it is expressed at all. We will return to this point below.

controversy, the declarative/nondeclarative classification has broadly enjoyed consensus and success.[93] It is generally accepted that these are served by distinct cognitive processes, and even by distinct neural processes.[94]

Furthermore, declarative memory has been divided into two categories. We sometimes merely remember that something is the case, as it is if one remembers that Samarkand is in Uzbekistan, or that David Hume died in 1776. We sometimes remember past episodes that we ourselves experienced, as it is in recalling one's first trip abroad, or recounting the birth of one's child. While it is true that we may also be able, conversely, to recall the episode of *learning* the geography of Samarkand, or to remember *that* one went abroad, these superficial similarities are generally not taken to pose a threat to the underlying distinction between remembering facts and remembering episodes. These two distinct ways of remembering were often blurred, even in experimentation in the middle of the 20th century, but in the last few decades their separation has enjoyed broad consensus among memory researchers.

The taxonomy of memory is thus one domain that has already benefited from conceptual analysis—many memory scientists explicitly mention the work of philosophers when introducing this classification of memory systems. The terms and categories for describing types of remembering have settled into something like cross-disciplinary equilibrium with these three classifications (or *memory systems*): episodic memory, procedural memory, and semantic memory.[95] Episodic memory may be the variety of memory that most often comes to mind first

---

[93] Tulving writes, "I shall assume the separation between propositional [declarative] and procedural [nondeclarative] memory as a fundamental given, although I can imagine how some psychologists may wish to disagree with the assumption" (Tulving 1983, p. 33).

[94] Dudai 2004, p. 60. Variations on this theme will be explored below.

[95] Not that this equilibrium implies consensus, or very much stability. The dividing line between episodic and semantic is contentious (see, e.g., Dokic 2001), as is the division between episodic and procedural (see, e.g., Mayes

when we consider memory; it is the ability to recall past episodes, to bring them to mind. As will be discussed below, *consciousness* is taken to be specially related to episodic memory, and researchers disagree as to whether any non-human cognitive systems can episodically remember. Procedural memory is the ability to remember how to do things—remembering how to play a piece on the piano, or remembering how to find the way through a maze. Procedural memory is often studied in non-human animals, and it bears close (but controversial) relations to learning and conditioning.[96] Semantic memory is being able to remember facts or the meanings of words. Remembering that Trieste is in Italy or that your birthday is in September are examples of semantic memory. Importantly, semantic and episodic memory are taken to be declarative in nature—expressible in words—while procedural memory is taken to be non-declarative.[97] In at least one sense this list—episodic, procedural, semantic—is often taken to be exhaustive. Although there are other distinctions to be made in explaining remembering, there are no other popular candidates for systems that stand side by side with these three.[98]

One last distinction worth mentioning, especially since it has garnered recent attention in philosophy, is that between short-term or working memory and long-term memory (what was once

2001). Whether "episodic memory", which originated with Tulving (Tulving 1983), even picks out a useful category is also questioned (Hoerl and McCormack 2001, p. 22). The taxonomy as a whole is sometimes challenged (Bussey, 2004; Bernecker 2009, §1.1).

[96] Procedural memory is sometimes conflated with implicit memory—the ability to remember something without awareness—but most researchers agree that these do not overlap perfectly, since priming experiments seem to show non-aware memory for facts, meanings, and episodes.

[97] This is relatively uncontroversial in the case of semantic memory. Controversy concerning episodic remembering that is inexpressible in language will be discussed below.

[98] The only candidate for a type of remembering that is sometimes mentioned alongside these is *autobiographical* memory—memory concerning one's self and past. The most common conception of autobiographical memory, however, is that it is a useful construct composed of elements of both episodic and semantic memory.

called short-term memory now tends to go by the name *working* memory).[99] The way that this distinction cuts across the three memory systems just described is controversial and often ambiguous. Even though it is assumed that working memory has its own dedicated neural process or processes,[100] and to some extent it is also assumed that long-term memory has its own dedicated neural system,[101] these systems are not side by side ways in which we remember, but rather they support different stages of the remembering process. The distinction between working and long-term memory is taken to hold at least in the case of episodic and semantic remembering—although it is most studied in semantic memory—and may have analogs in the procedural memory system.

2.2.2 *Encoding, Storage, Retrieval*

Although the details can vary, memory researchers, by and large, characterize memory in terms of a three-part process. This model divides the process of remembering into the following stages: the stage at which the remembered event, episode, fact, or skill is encoded into the brain or mind of the subject, the stage during which the resulting encoded memory or memory trace is inactive, and the stage at which the memory is retrieved and the past event, episode, fact, or skill is recalled or reproduced. Henry Roediger attempts to put this in relatively neutral terms thus, in a recent survey of the conceptual underpinnings of memory research:

---

[99] Working memory does not enjoy the consensus characterization one would guess it did from the treatment it is given by philosophers, but a good and influential starting point for surveying models of working memory is the work of Alan Baddeley (Baddeley 1992; 2003). Not all memory scientists condone the synonymy of working and short-term memory (Gathercole 2007).

[100] D'Esposito et al. 1995.

[101] Questions about just how long-term memory is neurally underwritten are plagued by empirical and conceptual difficulties, as will be demonstrated below, but an influential treatment of this hypothesis can be found in Lisman and Grace 2005.

> The learning and memory process can be conceptualized in three overlapping stages of encoding (also called acquisition, or learning), storage (retention, or persistence over time) and retrieval (access to, reactivation of or reconstruction of internal traces).[102]

Roediger is explicitly courting alternate conceptual possibilities in this piece for reasons I will return to below—traditionally, memory has been straightforwardly modeled on the stages of encoding, storage, and retrieval.[103] That is, at the time of an experience, the relevant input is *encoded* in the brain—and this is sometimes characterized literally, in terms of a neural code—where it remains as a coded, stored item until it is retrieved. Like the "memory" of the digital computer, there is an isomorphic relationship between the physical, encoded items and the experiences of times past. The memory traces which result from the encoding process would allow past experiences to simply be "read off" of them with just the right decoding mechanism, and it is something like this that happens at retrieval. The information in an experience at one time can only be re-presented at a later time by virtue of the coded trace connecting the two.

This basic encoding-storage-retrieval account may be most easily understood in terms of good old-fashioned episodic memory, for which it is a past, experienced episode which is encoded in the trace. The trace then persists in a stored form, relatively stable and relatively inactive, until it is activated by a retrieval process. This retrieval process ends with the past episode being brought to mind in the present by the remembering subject. Although episodic remembering is an easy paradigm case, the encoding-storage-retrieval account is applied broadly, across memory systems. Memory scientists often talk of semantic memory—or, to a somewhat lesser extent, procedural memory—being stored and retrieved as well. The human rememberer undergoes some

---

[102] Roediger 2007, p. 207.

[103] It is worth adding here that a fourth phase, consolidation, is sometimes places between encoding and storage. The consolidation of new memory traces into efficiently stored items seems, at least on the face of it, to be no less computer-inspired. Consolidation will be further considered below.

experience—be it the learning of a fact, experiencing of an event, or the training of a skill—and then stores the memory of this experience until it is needed, at which point the memory is retrieved and the subject can use and express the past information in various ways.

To be sure, the above is a simplified and overtly computer-like description which perhaps no memory scientist would, today, explicitly accept in its entirety.[104] In fact, a significant component of the present project will be to survey the various criticisms that modern memory researchers themselves have made of these models, in order to better understand just where such models have gone wrong. Some of these criticisms and concerns have been long standing: the renowned psychologist Frederic Bartlett was documenting the constructive, dynamic, non-archival nature of the remembering process long before anyone was taking the digital computer as a guide to the workings of the human mind.[105] Despite the fact that Bartlett's ideas fell out of favor during the cognitive revolution of the middle twentieth century, all recent memory researchers are quick to defend the constructive and dynamic nature of memory that Bartlett sought to demonstrate. In general, research and models which emphasize these features have been on the rise in memory science for decades. Nonetheless, this storage paradigm continues to be the foundation for memory science, cognitive science which works with memory, and for our understanding of human remembering more generally—despite the cracks in the edifice which have become apparent through the empirical work of the last several decades.

---

[104] Although Gallistel and King's recent (2011) book is a defense of exactly this. In the preface, their stated purpose is to articulate and defend the following simple message. "There must be an addressable read/write memory mechanism in brains that encodes information received by the brain into symbols (writes), locates the information when needed (addresses), and transports it to computational machinery that makes productive use of the information (reads)." Gallistel and King go on to remark, "such a memory mechanism is indispensable in powerful computing devices, and the behavioral data imply that brains are powerful organs of computation." Although the explicitly hypercomputationalist attitude evinced by Gallistel and King is a perhaps uncommon in this century, their basic assumptions are not.

[105] Bartlett 1932/1995.

It is not difficult to find recent psychologists of memory admitting that the basic assumption that memory is a matter of storage and retrieval has not changed. Jens Brockmeier recounts some of the changes that the sciences of memory have seen in the last decades, but laments the fact that these underlying conceptual frameworks have lingered.

> What has not changed is the hypothesis that each memory system operates according to the model of storage, that is, 'encoding,' 'storing,' and 'retrieval' (or 'recognition) of information, as a look in any psychological and neuroscientific textbook or research publication on memory testifies.[106]

Perhaps it is not surprising that these conceptual foundations have proven difficult to revise. Memory is studied by researchers with a diverse collection of disciplinary backgrounds, and what unity has been supplied by the computer metaphor and folk intuitions has probably been favorably received, and to a great extent helpful. John Sutton notes the inherent multidisciplinarity of the study of memory, and the difficulty this causes for revising the traditional view:

> Not all theories have taken memory to be a place where dead parts of the past sit passive until recalled to full presence. But, across the bewildering range of disciplines in which models of memory are constructed and criticized, vast gulfs between brains and society…limit moves beyond the archive.107

Even within disciplines, however, moving "beyond the archive" is no straightforward matter. The language and concepts in which the studies of memory are couched define the character of the problem spaces being investigated. Hence, some problems of memory science are deeply entwined with the encoding-storage-retrieval picture. It is common in the case of various amnesias to try to determine whether these are failures of encoding or failures of retrieval (and in some cases one or the other is taken to have been demonstrated, or at least made probable). Cognitive neuroscientists

---

[106] Brockmeier 2010, p.9.

[107] Sutton 1998, p.2.

seek the molecular and chemical basis for the storage of memory traces, and the action potentials associated with memory encoding.

Nor are scientists alone in systematically assuming that human memory is fundamentally a matter of storage. Philosophers would not talk about "memory files", or use notebooks as analogies for memories, were this the case. From memory-swapping thought experiments in work on personal identity to memory-implanting thought experiments in work on skepticism, philosophical discourse is often also built on the assumption that stored memory traces are, at least in principle, separable from respective encoding and retrieval processes. The entry of the computer onto the scene came just in time for the "cognitive revolution", and work in the philosophy of mind since the mid-twentieth century has been just as susceptible to excessive use of computer metaphor as work in the sciences has been. As in the sciences, the assumptions about the properties of the memory trace here have often—though certainly not always—gone unexamined. It is worth, in brief, to lay out some of these properties here.

2.2.3 *Engrams and traces*

In the descriptions above, I sometimes described that which was stored and retrieved as the "memory" itself. This locution gets us started in modeling memory, but of course if it is memory which we are taking ourselves to give an account of, then we cannot invoke "memory" as a part of our explanation. Characterizing this subject of storage and retrieval as a memory "trace" is one way to talk about that which is actually encoded, stored, and retrieved while avoiding circularity or strong theoretical commitments. In fact, one definition of 'memory trace' popular among memory scientists is that of Endel Tulving, who defines it simply as the sum physical change resulting from an event which is necessary for future remembering of that event, whatever these changes amount to:

> A memory trace is the neural change that accompanies a mental experience at one time (time 1) whose retention, modified or otherwise, allows the individual later (at time 2) to have mental experiences of the kind that would not have been possible in the absence of the trace.[108]

This notion of the trace is deliberately empty of empirical content, leaving the details to be filled in by future empirical work. It is noteworthy, though, that Tulving's trace is *neural* rather than *cognitive*. This is not always a feature of the trace as it is characterized by memory scientists, and the relevant distinction is not always made clear. The memory trace, as it is most commonly accepted in memory science, is a sort of functional placeholder—it is usually neural, but sometimes even this aspect is ill defined. The memory trace nevertheless plays a key functional role in explanations of remembering, as is demonstrated in this characteristic invocation by Yadin Dudai:

> The long-term trace persists in a dormant, inactive, representational state until reactivated in retrieval—or possibly also, in a behaviorally opaque manner, in the course of reorganization or maintenance of memory systems. The activation of the trace implies re-creation or reconstruction of a certain spatiotemporal pattern of neuronal activity. The distinction between inactive and active states of the trace deserves more attention than it has so far received.[109]

As we have already seen in chapter one, Dudai is among those memory researchers who has explicitly demonstrated sensitivity to conceptual and philosophical difficulties that arise in the language and concepts of memory research. It is also worth noting that he is pre-eminent among specialists in the neurobiology of memory, and that the above passage is one of the concluding and

---

[108] This is a relatively recent and updated version from Tulving (Tulving 2007, p. 66), but it develops on influential themes in Tulving's earlier work (e.g. Tulving 1983). Compare to Schacter's engrams, which are "the transient or enduring changes in our brains that result from encoding an experience" (Schacter 2008, p. 58).

[109] Dudai 2004, p. 79.

synthesizing remarks in a very influential and fairly recent article on the neurobiology of memory consolidation. Dudai's characterization of the role of the trace is representative of the character it is given by many in the field. Rather than succeeding in being empirically innocuous, it will be my contention that the picture of the traces employed by various scientists and philosophers has been all too informed by the computer metaphor. As an encoded, stored, and retrieved item, various characteristics are imputed to the trace which will ultimately face empirical and conceptual difficulties.

2.3 Computationalist assumptions in traditional memory models

In his book *Being There: Putting Brain, Body, and World Together Again*, Andy Clark describes "memory as retrieval from a stored, symbolic database" as one of the key characteristics of classical, computer-inspired cognitivism.[110] Although Clark admits that cognitive science has proceeded beyond this classical view in many ways, he also maintains that some of its tenets linger. In one section of the book entitled "After the Filing Cabinet," Clark credits connectionism with teaching us valuable lessons about the failures of storage, but also worries that these lessons have not taken full effect.

> Work with artificial neural networks thus provides a valuable antidote to what has been termed the "filing cabinet" view of mind: the image of mind as storehouse of passive language-like symbols waiting to be retrieved and manipulated by a kind of neural central processing unit. Nonetheless, some residual features of the filing-cabinet view remained unexpunged.[111]

[110] Clark 1998, p. 83. Clark cites this as one of five defining characteristics, the others being problem solving as inference, centralized cognition, the environment as a problem domain, and the body as an input device. Stored, symbolic memory is the first of these.

[111] Clark 1998, p. 67.

It is certainly true that the successes of connectionism in the mid-1980's and since have helped to undermine the argument that computationalism was the only game in town. It is also true that neural networks have helped to offer alternative views of learning, retained capacities, and re-presentation of past events. I will argue, though, that Clark's characterization of the "residual features" leftover from the dominance of the computer metaphor is dramatically understated. Some of the most basic tenets of the framework used to characterize memory are features of the antiquated storehouse view.

The compound assumption that memory must be supported by a collection of cognitive and neural processes that store and retrieve an encoded memory trace could be decomposed in many ways. It will be useful to divide this underlying framework into three basic assumptions which, although interdependent in some ways, can be investigated singularly. First is an assumption of modular or componential explanation. That is, the researcher who seeks to explain a memory process in terms of the parts of that process is assuming that what it is to explain remembering is to explain the component processes of remembering (and, in turn, the component phases of these component processes). This style of explanation is effective when explaining the computational processes found in, for example, a digital computer, and it lies near the foundations of many mainstream accounts of memory.

The second assumption is that there is a memory trace which is in some way structurally analogous to the remembered event or experience. Just what form this trace takes is a controversial matter, as is the role it plays in the processes that are posited to explain memory. Nevertheless, the basic assumption that some form of trace carries the information from past event to future remembering is nearly ubiquitous. Once more, the encoded physical trace that stores information is an excellent way to understand the "memory" of a digital computer, and once more this assumption as it is found in the study of human memory deserves careful philosophical attention.

The last basic, computationalist assumption about memory which pervades current models is the belief that the phenomena of memory are best explained in terms of hidden processes. This belief is importantly bifurcated. On one level it is claimed that there is a hidden cognitive process which is the process of remembering—an inner process akin to software-level processes in a digital computer. On another level it is claimed that there are hidden neural processes which underlie the phenomena of memory—inner biological processes akin to hardware-level processes in a digital computer. These claims are often confounded, and talk of inner processes of memory does not always clearly distinguish between the two. Furthermore, there is a spectrum of available claims for each of these, varying from weaker to stronger. In their weakest versions, these assumptions about inner processes are trivially true. In their stronger versions, they risk falsehood or unintelligibility. Once more, distinctions are not always clearly made when inner processes are invoked.

It will be my contention that these assumptions are impediments to clarity and progress in the science of memory. This is not to make a claim about the value of computationalism in the history of cognitive science: perhaps conceiving of memory in computationalist terms was needed. Near the opening of *Cognition in the Wild*—a project that shares some ambitions with the current project—Edwin Hutchins writes that cognitive anthropology adopted certain assumptions as a result of the cognitive revolution. He then writes:

> Perhaps these narrowing assumptions were necessary to get the project of cognitive anthropology off the ground. I will argue that, now that we are underway as a discipline, we should revoke these assumptions. They have become a burden, and they prevent us from seeing the nature of human cognition.[112]

[112] Hutchins 1995, p. xii.

Replacing 'cognitive anthropology' with 'memory science' makes this sentiment no less apt. Now that memory science is underway as a discipline, these computationalist assumptions should be revoked.

2.3.1 *Independence and modularity*

The first assumption which deserves our attention is that the mental states and processes of remembering are robustly and meaningfully modular—the assumption that any adequate explanation of the processes of remembering must decompose the process into its modular components, and in turn explain these components. This requires that the component processes and phases enjoy at least a modicum of conceptual independence from one another. This approach to the explanation of mental states and memory processes has become entrenched under the influence of the computer metaphor, but it did not begin there. Controversy surrounding the modularity of the processes of mind has a long history. William James offers this very assumption as one of the most nefarious in the study of mind. His articulation of this concern, so close in spirit with much of the present work, is worth quoting at length:

> The fundamental conceptions of psychology are practically very clear to us, but theoretically they are very confused, and one easily makes the obscurest assumptions in this science without realizing, until challenged, what internal difficulties they involve. When these assumptions have once established themselves (as they have a way of doing in our very descriptions of the phenomenal facts) it is almost impossible to get rid of them afterwards or to make any one see that they are not essential features of the subject. The only way to prevent this disaster is to scrutinize them beforehand and make them give an articulate account of themselves before letting them pass. One of the obscurest of the assumptions of which I speak is the assumption that our mental states are composite in structure, made up of smaller states conjoined. This hypothesis has outward advantages which make it almost irresistibly attractive to the intellect, and yet it is inwardly quite unintelligible.[113]

[113] James 1890/2007, p. 145.

James' concerns did not carry the day; researchers have by and large accepted—both tacitly and explicitly—this assumption. Cognitive psychologists who study memory argue about whether there is, for example, a consolidation phase after the encoding phase but before the storage phase, but they almost never argue about whether this notion of component phases is the right framework within which to understand remembering. As James points out, the assumption that a process like remembering is composite in structure is almost irresistible. Explaining remembering is just taken to mean explaining the component processes and parts of the processes which constitute remembering. This goal, to explain memory by isolating the component processes which comprise it, aims at what Andy Clark calls componential explanation.

> To explain the functioning of a complex whole by detailing the individual roles and the overall organization of its parts is to engage in componential explanation. This is the natural explanatory style to adopt when, for example, we explain the workings of a car, a television set, or a washing machine. We explain the capacities of the overall system by adverting to the capacities and roles of its components, and the way they interrelate.[114]

Clark is right to claim that this approach is all too common in explanations of cognitive phenomena. Clark claims that although this manner of explanation has its virtues, it is not the right way to go about explaining cognitive systems. Part of the reason for this, according to Clark, is that componential explanation usually depends on a robust and central role for representation.

> In short: there is a relation between the componential analysis of intelligent systems and the image of such systems as trading in internal representations, for the distinctive roles of the posited components are usually defined by reference to the form of content of the internal representations they process.[115]

Clark may be right that this one reason which motivates the ubiquity of the componential approach to mental processes and systems. And in fact, Jesse Prinz makes a compact case for componential

---

[114] Clark 1998, p. 104.

[115] *Ibid*., p. 105.

explanation of cognitive systems, based on the fact that representations are the currency of these systems:

> Most cognitive scientists agree that the mind is a computer, in some sense. It is a device that processes information by transforming representations in accordance with rules. Computational devices decompose into various interconnected subsystems, each of which performs some aspect of a complex task.[116]

Still, researchers' explanation strategies must also be driven by the apparent and widespread success of componential explanations for a broad range of non-representational, non-intelligent phenomena. Successful explanations are not always componential. Only in rare case does the explanation of a word involve explaining the word's parts, and it is only in spurious fashion that the explanation of a chess-piece involves explaining the parts of the piece. However, componential approaches to explanation are, in a wide range of cases, the first and best methods to come to mind. This may be good reason to accept componentiality provisionally, but it does not provide an adequate defense of these assumptions when questioned.

In the study of memory, claims of componential explanation can be usefully divided into two genera. First there is the claim that what we call remembering is constituted by several distinct, alternative, independently explicable, processes. Daniel Schacter sums up the different kinds of memory in his recent comprehensive work on memory science thus:

> Most important, we have now come to believe that memory is not a single or unitary faculty of mind, as was long assumed. Instead, it is composed of a variety of *distinct and dissociable* processes and systems. Each system depends on a particular constellation of networks in the brain that involve different neural structures, each of which plays a highly specialized role within the system.[117]

These processes have different functions and different neural and cognitive bases, but happen to be subsumed under our everyday concept of memory. This assumption of modular process types

---

[116] Prinz 2012, p. 49.

[117] Schacter 2008, p. 5, emphasis added.

is what drives the memory systems divisions introduced at the beginning of this chapter. We might depict the remembering "process" visually such that these divisions would run horizontally, as it were. That is, this variety of decomposition posits parallel, alternative subprocesses which, taken together, just are remembering. For the sake of convenience, let us refer to the assumption that the remembering process is composed of modular, alternative, subprocesses which run parallel *horizontal componentiality*.[118]

In addition to this, there is also the assumption that the remembering process, or any of these subprocesses that may have been decomposed from it horizontally, is composed of component processes or phases which can be decomposed *vertically*. That is, imagining these horizontal processes of remembering, each or all of them may be constituted by independently explicable subprocesses which lie end-to-end and must be decomposed by dividing these processes vertically. Let us refer to this assumption as *vertical componentiality*. As such, a model of memory which divides remembering into types of remembering, and then which seeks to, for example, find the neural process which correlates to semantic memory as opposed to episodic memory, is committed to the horizontal componentiality of memory. A model of memory which divides remembering into a chain of conceptually independent phases such as encoding, storage, and retrieval, is committed to the vertical componentiality of memory. We have already seen some evidence that most of the mainstream accounts of memory are committed to both of these componentiality assumptions: any model that attempts to explain the memory process by explaining the component phases of encoding and retrieval is assuming that the remembering process can be fruitfully decomposed vertically, and any model that posits different memory

[118] In psychology and neuroscience, countercurrents that emphasize the non-decompositional nature of the brain (and mind) run at least as far back as the influential work of Karl Lashley (1929), and, as we will see, to William James (1890). Nonetheless, functional decomposition into distinct systems has been the rule rather than the exception for much of the last century.

system processes, which independently underlie different remembering types is assuming that the remembering process can be decomposed horizontally.

There are some processes which do not lend themselves to this kind of decomposition. This must be so, of course, if we are to avoid infinite regress: process decomposition must bottom out somewhere. However, there are even plenty of standard, high level processes which are so constituted. Nobody would advocate understanding the process of checkmating an opponent in terms of the phases of checkmate (although one could certainly decompose it horizontally into the different types of forms checkmating can take), and it is unhelpful to try to explain the game of chess by explaining the independent forms chess games can take (although one will decompose the game vertically into phases as part of the explanation). There are probably many cases that lie near the boundaries of useful componential explanation, but what matters is that this style of explanation can be more or less appropriate depending on the explanandum. This means that the assumptions of componentiality are answers to the question of whether remembering is an appropriate subject for decomposition—and this question is very infrequently asked in any explicit terms.

Explaining some phenomena in terms of the explanation of its parts is sometimes called *mechanistic* explanation. The family of hypotheses which attempt to clarify mechanism, mechanistic explanation, and the role of these in mental explanations has generated a broad and variegated array of literature in the philosophy of science, philosophy of mind, and cognitive science.[119] Many of the disputed theses therein are orthogonal to, but merely orthogonal to, our

---

[119] William Bechtel, whose work has been an influential starting point for discussion of mechanism and explanation, has recently argued that mechanism does not require strong modularity (Bechtel 2009); he accomplishes the reconciliation of mechanistic explanation and non-modularity by treating psychological mechanisms as "nearly decomposable", separable but not isolable. This strategy is akin to what will be taken up below.

present concern. No results concerning the way that an assumption of componential explanation shapes the study of human memory will provide direct evidence for or against the general plausibility of mechanistic explanation, nor will they illuminate the nature of this style of explanation in a decisive manner. I take it that there are no parties to any of these debates who argue that mechanistic or componential explanation can never be helpful. For example, explaining an internal combustion engine or how a search algorithm accomplishes its goal is bound to be a mechanistic enterprise. Conversely, I also take it that there are no parties who argue that a mechanistic or componential explanation is the only explanation available for every phenomenon. Explaining what a knight is in chess, or what an apex predator is in ecology, will very likely not include any discussion of their respective parts. Our present interest is in better understanding the role of this assumption in accounts of memory, demonstrating some of the difficulties this has produced, and trying our best to understand what modeling memory might look like in the absence of such an assumption.

2.3.2 *Isomorphism and the memory trace*

Although Tulving's definition of the trace, strictly speaking, does not mention any properties which are structural analogs of the remembered and encoded event, the very concept of *encoding* guarantees such isomorphism.[120] Most memory researchers understand this implicitly, and do not hesitate to characterize the trace in a way that bespeaks an isomorphic relationship between at least some part of the remembered features of the past event and some part of the physical structure of the memory trace. Any model that posits a stored, encoded, localizable *thing,* which is a result of a past event, and which causally allows for remembered information to later

[120] Strictly speaking, these isomorphic mappings between parts of structures might be more accurately called homomorphisms, which are more general than the (bijective) isomorphisms that they may also be, but I will follow most cognitive science authors in what seems like a benign casual usage of the term 'isomorphism'.

be recalled, must also claim that this information exists in the *thing* itself, albeit perhaps in an encrypted mode. Indeed, most researchers go further and do not hesitate to talk of the storage and retrieval of traces that are representations. Yadin Dudai notes that the "question of how information is coded and represented in brain and cognition" is considered by many researchers to be "the most crucial problem in the neurosciences."[121] For any representation of something that allows for any features of the represented thing to be read off from the representation alone, there must exhibit some structural analog between representation and that which is represented. John Sutton writes that "this idea that a 'trace' acquired in past experience somehow 'represents' that experience, or carries information about it, is at the heart of …the dominant view of memory in modern philosophy of mind, and it is assumed in much work on memory in cognitive science."[122]

Although ubiquitous and fundamental, the assumption that the memory must utilize a trace is perhaps the assumption which most lacks a singular and explicit articulation. That some kind of trace is almost always invoked in memory research is an easy claim, but pinning down the nature of this trace in any fashion that faithfully represents the views of "most researchers" is probably impossible. Part of the reason for this, of course, is that, as Susan Sara points out, "the 'trace' can never be accessed directly. Strictly speaking, it is a hypothetical construct whose very existence is inferred from its retrieval and behavioral expression at the retention test."[123] This leads researchers to a spectrum of commitments concerning the memory trace. On the spare end of the commitments spectrum, we have Tulving's "neural change." Additionally, many researchers

---

[121] Dudai, in Roediger et al. 2007, p. 53. The fact that these are characterized as representations is probably not as important as the role they play as traces, at least insofar as problems and resolutions in the science of memory go. This will be explained further in the following section.

[122] Sutton 2012, §2.

[123] Sara 2007, p. 184.

impute isomorphism and representation to the trace without argument or hesitation. Some researchers treat the memory "trace" as both cognitive and neural, or as alternatingly cognitive and neural.[124] Many researchers substitute other concepts and words in for what it is that is encoded, stored, and retrieved (neural firing patterns, changes in behavior, neural representations, learned information, memory representations, experience, and "brain changes", just to name a few). Although the conflation of cognitive and neural can seem bewildering, one way to understand the common conception and how neural traces are the subject of cognitive processes is just to take the digital computer as our guide. Encoding is seen as analogous to a software process that is "saving information" into the computer's memory. Retrieval is seen as analogous to a software process that is locating and utilizing the stored information in the computer's memory. Note that there are parallel hardware processes which underlie these software processes, of course, but also that there is no parallel software process which lies atop the hardware process of storing the "memory." During the storage phase, the hardware is, at least more or less, doing the whole job. It is in this way that models of memory can get away with parallel cognitive/neural process pairs for the first and last phases while relying only on neural traces for the storage phase. The cognitive "storage phase" is not just undergirded by a certain neural process—it is exhausted by the neural.

The same cannot be unequivocally said of encoding and retrieval. Cognitive psychologists work on modeling aspects of encoding and retrieval at the *cognitive* level, expecting that neuroscience will reveal the physical processes that support these. Neuroscientists work on identifying *neural* processes that are active during encoding and retrieval, with hopes that these

[124] This situation seems to have been improving in the last few decades—as will be elaborated below, where researchers were once very frequently committed to traces as "ontologically ambiguous mental items", as Sutton writes, the view that the trace is neural and that the only cognitive correlate is the non-process process of storage has gained wide acceptance (Sutton 2012 §2.1). Still the confusion remains—Sara seems to be writing of a "cognitive" trace in the remark cited above.

will fit together with the still amorphous cognitive processes. For these reasons, the level at which encoding or retrieval is being studied is often made more explicit than the level at which the trace is being studied. Nonetheless, the memory trace is a crucial component to most models of memory, and whether or not it is representational in nature the trace is almost always characterized as isomorphic to (at least some aspects of) that which is remembered. That is, the trace is taken to explain the resemblance between the past event and the present remembering—these shared features were stored, as it were, in the trace. Therefore the trace must contain whatever information is preserved, even if this contained information is encoded. The supposition that memory demands an isomorphic neural trace of some kind is one common to many models of memory, and it is also one that derives, at least in part, from the computationalism latent in memory science.

> *The trace isomorphism assumption*: There exists a neural trace of any remembered experience which is structurally isomorphic, in some of its features, to some features of the past event or experience.

It is worth immediately noting two things about the way this assumption is at work in the science (and philosophy) of memory. First, the trace is also isomorphic to certain features of the later experience or phenomena of remembering. In fact, the trace is posited as that which explains the isomorphic relationship between some features of the past event and some features of the remembering process. That is, it is assumed that when the subject remembers a past event, the fact that the memory *resembles* the past event is explained by the fact that a record which captured certain features of the past event was made in the brain. An easy analog might be the way that an acoustic recording might be carved into a vinyl record. The fact that the music, when played back later, resembles the band that played into the horn of the gramophone is explained by the structure of the physical record and its isomorphic relationship to both instances of music.[125]

---

[125] It is for this reason that the gramophone is actually adopted as a model for memory not infrequently (Martin and Deutscher 1966, p. 189; Roediger 1980)

Second, the claim that the trace is structurally isomorphic to past and future events is often not explicit. In fact, we already saw that Tulving's definition of a trace does not include this feature, for reasons to which we shall eventually return. Ultimately, I will argue that the blank quality of Tulving's definition actually undermines the job that the trace is supposed to be doing in an archival model. At present, it is sufficient to note that most researchers treat the memory trace as though it were isomorphic, including Tulving himself.

> It is important to note that it is seldom possible for an experimenter to completely determine the informational contents of a memory trace. At best these contents can be influenced or biased by manipulating the conditions of encoding. Many features of events are encoded…[126]

Despite keeping feature-matching isomorphisms from the definition of trace, Tulving routinely depends on this matching when using the concept of the engram or trace. Many do; it is thought that the stored trace can only explain the preserved features of past experience if these features are encoded in it. The very concept of "encoding"—one of the primary and least controversial phases of the remembering process according to most models—implies that the resulting trace contains coded features which map to features of the past event or experience.

It is not hard to see the trace isomorphism assumption as an inheritance from the computer metaphor—this trace is just the manifestation of the "stored symbolic database" that Clark imputes to the foundations of computationalism about the mind. Furthermore, many of the specific alleged characteristics of the trace which have recently run afoul of empirical and conceptual considerations are just those which appear most computationalist. Taking computer storage as our guide, any given stored input is physically localizable within the hardware. That is, a trace that is stored is stored somewhere. Although cognitive neuroscientists have largely given up on tracking

[126] Tulving 1983, p. 250.

down so-called grandmother neurons—the specific cell or cells which represent your memories of and ability to recognize your grandmother—the notion of physically located memory traces has not been an easy one to dissolve. The assumption that traces for particular "memories" must be somehow localizable within the brain (even if we cannot presently locate them) is one that pervades much of the cognitive science of memory, despite the pressure on this assumption which has come from the study of neural networks in recent decades. Although there are theorists who explicitly eschew localized storage (but who still generally invoke an encoding-storage-retrieval model of memory),[127] one of the main properties of the trace that has gained acceptance through the rise of the computer metaphor is certainly this claim that the trace of a past event is stored in a definite and localized place in the brain.[128]

Another property that has often been tacitly imputed to the trace, and is in keeping with computer information storage, is the idea that the memory trace lies inert while stored. Once a memory trace is encoded, it lies dormant in its coded form until the proper (or improper) retrieval procedures are initiated. What it is for the encoded item to be in *storage* at all is just for that item to be inactive, even if, as Yadin Dudai writes above, "the distinction between inactive and active states of the trace deserves more attention than it has so far received."[129] A related property of traces that has been strengthened by comparisons with computers is the notion that a given memory trace defines the product of retrieval. That is, for any one encoded trace, there is just one memory "product" upon recall—the information in the trace is just the information in the past event, which

---

127 See Michaelian and Sutton 2013 for an example of this, as well as a review of distributed or nonlocal traces in general. These distributed traces will be discussed below.

128 Ramsey makes a similar point about representation, writing: "Indeed, researchers often skip the question of *whether* neural receptors function as representations, and instead ask about how the actual representational encoding is done" (Ramsey 2007, p. 120, original emphasis).

129 Dudai 2004, p. 79.

is again just the information presented at retrieval. Retrieval is essentially binary: either a memory has been retrieved, or it has not. Any changes in the product of retrieval are taken to be incidental or environmental: the encoded trace has a definite informational character which, though it may perhaps be obscured, defines what can be attained during retrieval. It makes sense to talk about the underlying, encoded, memory trace even in the absence of its successful retrieval. These characterizations are still considerably simplified for the sake of focus. As it happens, most memory researchers (and many cognitive scientists in general, as well as philosophers) will reject at least some of the foregoing descriptions of the stored memory trace.[130] Nonetheless, even in those models and accounts where these aspects of the storage approach are controverted, I will argue that the general conceptual framework is often kept.

Philosophers have shown interest in this notion of the isomorphic trace as well. Sven Bernecker begins his introduction to a study of the nature of memory thus, assuming without argument that successful remembering must invoke a stored, representational trace:

> Since to remember is to retain some previously acquired representation, the analysis of memory must include some provision for one's having had the representation in question in the past and for one's still possessing the representation.[131]

The structural isomorphism between past event and present brain state especially attracts the attention of those interested in the representational properties of the trace. Gerard O'Brien and Jon Opie call this structural isomorphism a "second-order resemblance."[132] This refers to the fact that while there are not likely to be any resemblances between the first-order properties of a brain

[130] Schacter writes, "the idea that there is a one-to-one correspondence between a bit of information stored away somewhere in our brain and the conscious experience of a memory that results from activating this bit of information is so intuitively compelling that it seems almost nonsensical to question it. Yet scientists who study memory and theorize about it are increasingly skeptical of this idea" (Schacter 2008, p. 71).

[131] Bernecker 2009, p. 5.

[132] O'Brien and Opie 2004.

state and that which it represents (that is, no property of the neural configuration representing, say, a face, actually resembles "freckledness"), there are likely to be (and O'Brien and Opie argue that there are) resemblances among the *second*-order properties. There as many eyes as there are ears on a face, and there are as many neural-configurations-representing-eyes as there are neural-configurations-representing ears in the neural representation of a face. Philosophers are especially interested in isomorphism, or second-order resemblance, because this is thought to be one component of representation. William Ramsey counts this kind of structural isomorphism as an important variety of representation; he uses the notion of what he calls "s-representation" to restrict what can and cannot be expected of mental representation.[133] Our present interests lie first with isomorphism itself, and only incidentally with representation; ultimately it is whether or how the trace is isomorphic that I seek to make explicit and to criticize. Before going on to characterize the last computationalist assumptions, however, the representational nature of memory deserves a brief treatment.

2.3.3 *Excursus*: *Memory and representation*

Explanations of memory, and especially those that invoke the encoding of memory traces, often couch these explanations in terms of representation.[134] In both the science and philosophy of memory, there are sometimes taken to be neural traces which are representations of past events, and there are sometimes taken to be mental representations. The very notion of representation,

---

[133] Ramsey 2007, Ch. 3. Ramsey contrasts this structural representation with varieties of "input-output" representations that more closely track the symbols manipulated in standard computation.

[134] Indeed, Alessandro Treves equates the two, after considering the possible differences in connotation between coding and representation: "Such overtones will not be given further attention here: coding and representation will be considered as fully equivalent concepts" (Treves 2007, p. 55). Although the relationship may fall short of Treves' equivalence—it is not clear that representation requires coding—there are good reasons to think that coding may imply representation.

and especially of mental representation, is one of the most important and embattled in the philosophy and sciences of mind, as Ramsey points out in the preface to his recent book, *Representation Reconsidered*.

> It has become almost a cliché to say that the most important explanatory posit today in cognitive research is the concept of representation. Like most clichés, it also happens to be true. … Yet despite all of this attention (or perhaps because of it), there is nothing even remotely like a consensus on the nature of mental representation.[135]

This could make it seem as though a proper analysis of memory must wait for the dust to settle from the turbulent discussions of representation. Insofar as I am able, however, I will set these debates to the side for the purposes of this study of memory. Giving due respect to an analysis of representation would take us too far afield, and it will be possible to remain neutral on most core questions about representation while making real progress toward a critique and alternative account of memory. Although researchers often characterize remembering in terms of representations often, both in terms of neural, representative, traces and in terms of mental representations, these are usually doing little work, *qua* representations, for the models in question.

The assumption from the last section, that there must be a neural configuration that is structurally analogous or isomorphic to the past event, deserves careful scrutiny. It is also true that these neural traces are often described as "representations" of past events. Our foremost concern, however, is simply whether there must be isomorphic neural configurations. Such isomorphisms, if they existed, would have certain at least some features indicative of representation (causal-informational features shared with tree rings, etc.), and may or may not be good candidates for full-blown representation; these are either identical or very similar to the notion of simulation-

[135] Ramsey 2007, p. xi.

representation employed by Cummins[136] and the s-representation of Ramsey.[137] However, whether and how these count, or do not count, as representational will have little bearing on our development of a better account of memory. Neuroscience has so far failed to actually identify any such neural isomorphism, and its status in cognitive science and philosophy is highly contested. The job of determining whether such things must exist, or can exist, and what role they play in an adequate account of memory is more urgent (and probably more immediately soluble) than whether, if they existed and were important to models of memory, they would be genuinely representational.[138]

Sometimes researchers also talk about memory in terms of inner, mental, representations. These tend to play a more important role in models and arguments simply as components of inner processes than they do as representations qua representations. The assumption that the processes of remembering can be usefully decomposed into modular parts has already been identified as one which deserves attention. In the section below we will examine assumptions about hidden mental processes. Once more, the assumption that remembering must involve an inner process which ranges over component parts is already problematic. Whether these parts of hidden processes would, if they existed or were important to models of memory, count as genuinely representational is of secondary concern.

All of that being said, a critique that demonstrates that these inner "representations" are not necessary for what we call remembering, and that a good account of remembering need not invoke this notion at all, is not one that has remained entirely neutral on the topic of mental representation.

---

[136] Cummins 1991.

[137] Ramsey 2007.

[138] Additionally, I take it that some advocates or opponents of representational theories of mind do not require this kind of neural representation of memory for their arguments, though this style of representation—so-called s-representation—is one that Ramsey finds more theoretically viable than others.

Some of the arguments below have analogs that have been used to argue against the need for representation generally, and many of the kindred movements in the philosophy of cognitive science that were surveyed in chapter one are, broadly speaking, anti-representational. Having a representational theory of mind is sometimes just taken to mean the same thing as having a computational theory of mind,[139] and it is precisely these computationalist assumptions which I am investigating, and ultimately undermining. Thus an argument against computationalist assumptions in one important cognitive capacity, memory, can easily find a home among, and perhaps can trade resources with, anti-representational theories of cognition. Nonetheless, it is important to remember that none of our considerations here speak against the possibility of mental representation in general,[140] nor against the claim that mental representation is a robust and crucial feature of other core aspects of cognition. Whether the mind will ultimately be rightly described in representational terms, or not, the same criticisms pertain to isomorphisms and traces.

2.3.4 *Hidden processes, cognitive and neural*

The last assumption, or family of assumptions, to identify in the way that memory is studied concerns hidden processes. As we have already seen, remembering is taken to be the "conscious tip of an unconscious structure", in Dreyfus' words. The hidden cognitive process which underlies memory is implicit in almost all of modern theorizing about memory. Additionally, it is assumed that there are hidden neural processes that underlie these cognitive processes (though these neural processes are substantially *less* hidden, since there are ways, at least in principle, to observe them). One of the factors that makes this collection of assumption especially pernicious is that there are

[139] Pitt 2013, p. 1.

[140] In fact many of the arguments presented below presuppose the possibility that we do *sometimes* engage in what might rightly be called mental representation of the past.

several different hidden-process assumptions at work in the cognitive science of memory, and these are similar and tacit enough that they lend themselves to conflation and equivocation.

Even the most overarching distinction—between cognitive and neural processes—is not always made clear. Memory scientists will sometimes blithely and repeatedly wander across this distinction, resulting in nearly unintelligible claims:

> For example, during retrieval of a long list of words, individual words or sequences of words in the list may be retrieved in each retrieval phase of the theta cycle (50-60 ms) and then buffered in neocortical circuits for later generation of the response. Additional hippocampal retrieval may be suppressed or ignored during the time needed to generate the response. During a long delay between words being generated at a behavioral level, there might be many hippocampal theta cycles retrieving words which were already generated behaviorally, but which are rejected by prefrontal mechanisms for selecting responses.[141]

Just to spell this out, the claim is that there may be a subprocess which consists in *hippocampal theta cycles retrieving words*. The hippocampal theta cycles, or rhythms, are simply periodic oscillation in local field potentials throughout the hippocampus. These oscillations are not always present, and their presence correlates in interesting ways with different types of activities at the organism's behavioral level. These correlations deserve more study. However, whatever theta cycles may or may not do (and it is already unclear whether oscillations "do" anything in this sense), they do not retrieve *words*. Mixing the cognitive and the neural haphazardly does not address the gap there between. Unfortunately, the dominant strategy deployed in order to address this gap is all too often merely this attempt to interweave words and concepts from each of the levels under investigation, and to then hope that the problem goes away. Purging problematic assumptions from the study of memory will not, by itself, solve the mind-body problem: understanding how to rightly characterize the relationship between the cognitive and the neural is

141 Hasselmo 2007, p. 125.

a difficulty which extends throughout much of cognitive science and philosophy. However, the way forward can only be hindered by this kind of slippage.

It is uncontroversial that there are correlations between brain states and behaviors, even if it is not always clear how to characterize these correlations (or their implications). Many cognitive neuroscientists are sensitive to concerns about localization and inference due to their recognition of the amount of interpretation required to even connect brain data to any identifiable behavior, a difficulty emphasized by Kurt Danziger:

> Even when there is good evidence for some correlation between well-defined task performance and elevated activity in an identifiable cerebral system there remains the problem of interpreting such a correlation. To use such a correlation as a basis for anything more than a tentative hypothesis one would have to rely on the logical fallacy of affirming the consequent: only if we are already convinced that our procedures have indeed isolated a modular psychological function and a bounded brain region, both of which are generalizable across individuals, can correlation of the two be regarded as demonstrating a localization of function. Supposing such a demonstration to be successful, the localized pattern of brain activity would constitute only a necessary, not a sufficient condition for the performance of a particular human action. [142]

These "modular psychological functions" correlated with the activation of particular brain regions in the cognitive neuroscience of memory are often hidden cognitive processes, such as encoding or retrieval, which only incidentally yield behavior-level phenomena which provide the basis for correlation. This means that identifying the cognitive process that is the target of a supposed correlation is no straightforward task. Any subject who is demonstrating the putative behavior-level consequences of the cognitive processes of remembering is also often *ipso facto* demonstrating behavior-level consequences of countless other cognitive phenomena. Someone who is recounting a past event is also speaking, storytelling, interpreting, following instructions, thinking, choosing, and imagining. In his 1998 review of the status of the cognitive neuroscience of human memory, John Gabrieli emphasizes this:

---

[142] Danziger 2008, p. 237.

> There is also a great deal of psychological interpretation involved in understanding the meaning of an activation, i.e. in specifying what mental process is signified by an activation. Most imaging studies report activations arising from the difference between two tasks. Such differences are not only open to a variety of interpretations but also are often confounded with factors such as task difficulty or trial duration.[143]

Neuroscientists are aware of these difficulties, and strive to mitigate them with clever experimental design and cautious data interpretation. Nonetheless, perfect experiments that track exactly one inner cognitive phenomenon are implausible, and some careful interpretation is always necessary.

In memory science, the gap between neural and cognitive levels of explanation is often recognized, but more often than not this recognition takes the form of optimism about future discoveries of translation or subsumption. That is, the careful memory researcher will admit that no conclusions ought yet be drawn that do not cautiously preserve the integrity and independence of these levels of explanation, but will also intimate that this divide will soon be bridged by a better understanding of the rules of translation:

> Moreover, modifications in synapses and nerve cells should not be simply equated with modifications in memory. Those are different levels of organization and function, where one level (cellular) is assumed to subserve, by virtue of its plasticity, the other (behavioral), *but the translation rules that govern this interaction are not yet known*. Further, it is premature to determine whether identified functional or morphological changes in synapses, which are correlated with, and obligatory for, the processes of learning and consolidation, are indeed causally related to the encoding and persistence of memory per se, as opposed to auxiliary processes that are required for the formation of the trace but do not embody it.[144]

In this warning, Dudai is calling for caution and conceptual clarity in work on the cognitive neurobiology of remembering. He warns his fellow researchers against too hastily assuming correlations in the absence of supporting data. Nonetheless, he implicitly expresses confidence in the general approach: the translation rules between neural and cognitive processes of remembering are not yet known, but these rules comprise one of the targets of scientific investigation. This

---

[143] Gabrieli 1998, p. 89.

[144] Dudai 2004, p. 79, emphasis added.

confidence reveals the basic assumption, shared by the vast majority of memory researchers, that human remembering is an inner, cognitive process that is supported or implemented by a neural process. In accord with the vertical componentiality assumption above, most researchers no longer believe that all instances of remembering are necessarily instances of the same cognitive or neural process. The received view is that there are probably different modular cognitive processes responsible for each of the types of remembering—this is part of the reason these ways of remembering are referred to as memory systems—and that each of these cognitive processes is supported or implemented by its respective neural process.[145] Careful researchers often admit that our classifications are not yet perfect, but they are optimistic about our ability to divide memory into a certain number of cognitive processes which lie atop of their respective neural processes.

It is important to note that these classifications and correlations are assumed to yield type-type correlations. That is, the weaker assumption is merely that any *token* instance of a remembering process is some token cognitive process, which is in turn implemented or realized by some token neural process. To deny the first part of this would be to deny that remembering is a cognitive process at all, and to deny the second part would be to deny that cognition is, in general terms, realized or implemented neurally. Researchers investigating the correlations between neural and cognitive processes are assuming a stronger connection than this. It is not just assumed that *this particular instance* of remembering is a cognitive process that is implemented neurally, but also that this instance belongs to a *type* of cognitive process that is implemented by a *type* of neural process. This distinction deserves our attention, for to deny that we have grounds to assume type-type correlations between cognitive and neural processes is sometimes—mistakenly—taken as a denial of the cognitive, or a denial that cognition is implemented by the brain. Even after

---

[145] See Schacter and Tulving 1994 for a thorough elaboration, or Squire 2004 for a more recent survey of this consensus and its development.

admitting that we engage in cognitive processes, and that these processes are neurally implemented, it is a further assumption that there are neural process types, which can be individuated, that track cognitive process types to which our instances of remembering belong.[146]

Consider examples of cognitive processes that do not belong to a cognitive type that can be correlated with a neural process type. For example, take joking, expecting, or promising. Any token joking process is also a token cognitive process. However, it seems obviously misguided to look for the cognitive process type to which all (or even many) joking instances belong. This is no denial that joking is a cognitive activity, it is only the denial that there is a cognitive activity that can be typed that *is* joking. Similarly, we do not imagine that we can find a neural correlate of promising—even though surely every instance of a promise is a token of cognition that is in turn (at least in large part) implemented neurally. Ryle made this point concerning belief and knowledge, cautioning philosophers not to "fall into the trap of expecting dispositions to have uniform exercises," or "one-pattern intellectual processes in which these cognitive dispositions are actualized."[147] Memory is not just taken to be cognitive and neural in this manner. Rather, it is assumed that any given remembering instance belongs to a cognitive process type that is in turn implemented by a neural process type. Once more, this assumption is clearly consistent with an approach that looks to the digital computer as a guide to cognition: an instance of computer output really is evidence for a particular software-level process, which is implemented by a particular type of hardware process. Thus along with assumptions of vertical and horizontal componentiality,

---

[146] Robert Cummins writes, "Everyone who is not a dualist believes that mental processes are processes that go on in the brain" (Cummins 2000, p. 133). Even setting worries about embodiment and situatedness aside, this sentiment, though common, is problematically ambiguous: admitting that there are no non-physical components to a mental process does not imply the stronger reading that there is some identifiable neural process that is identical to the given, identifiable mental process.

[147] Ryle 1949/2009, p. 32.

and the assumption that memory depends on a neural trace that is structurally isomorphic to the past event, the last computationalist assumption in the science of memory is one of hidden processes.

> *The hidden processes assumption:* Any given instance of remembering is subsumed under an identifiable cognitive process type that is implemented by an identifiable neural process type.

It is uncontroversial that the details of these processes, and the correlations between them, elude easy characterization. Part of the difficulty in settling on a cognitive model of the remembering process or processes is simply due to the fact that no corresponding neural processes have been established. The failure to identify brain processes which are the implementation of the cognitive processes of remembering has even pushed some researchers' investigations *down* a level, to seek the underlying processes at the molecular level:

> There must be special processes that underlie stable structures and these must be explainable in molecular terms. In summary, I argue that there must be memory molecules that have a special ability for stable information storage. … My view is that once memory molecules responsible for persistence are identified, the investigation of persistence can be put on a solid footing. It will be possible to separate persistence from retrieval processes by direct biochemical tests.[148]

It is perhaps no surprise that these processes have eluded discovery. In fact, as they are characterized, the sort of neural or molecular process which is needed has become something quite mysterious, as can be gleaned in this description by David Sweatt:

> Overall then, I conclude that molecular mechanisms involved in retrieval must be able to impinge directly upon the molecular basis of the engram, simultaneously triggering a re-activation of many of the same molecular changes involved in establishing the engram initially. Something very interesting is happening at the molecular level with retrieval—there seems to exist a specific mechanism capable of halting or erasing a chain reaction, but re-starting it in a plastic form that may end up being slightly different. A mechanism that is capable of accomplishing this is quite mysterious to me based on present knowledge.

148 Lisman 2007, p. 206.

> Nevertheless, having hopefully deduced its existence, the mechanism should be amenable to experimental investigation in the near future.[149]

While not all those who study the cognitive neuroscience of memory are as sanguine about this near-term possibility as Sweatt, he is no exception in expecting that sorting out the right cognitive processes and the right correlated neural processes will eventually result in an explanation of memory. Sweatt is also not alone in moving quickly past any defense or even explicit statement of the assumption that remembering is supported by certain cognitive process types, which are in turn implemented by certain neurobiological process types. These assumptions lay groundwork in the models for subprocesses like encoding or retrieval, which are inner cognitive processes that are implemented neurally. Whether such inner processes of remembering exist at all, or whether these are amenable to neural correlates, is rarely asked.

Since many researchers who study memory, in both philosophy and the sciences, admit that some of the difficulties memory studies have encountered are in part due to conceptual issues, it can only aid matters to make these assumptions explicit. In so doing it will ultimately become clear that this underlying computationalist framework is responsible for detrimental confusions in the studies of memory.

---

149 Sweatt 2007, p. 212.

CHAPTER III

EMPIRICAL DIFFICULTIES FOR ARCHIVAL MODELS

> It is astonishing what havoc is wrought in psychology by admitting at the outset apparently innocent suppositions, that nevertheless contain a flaw. The bad consequences develop themselves later on, and are irremediable, being woven through the whole texture of the work.
>
> William James[150]

> I believe that every first-rate cognitive neuroscience laboratory now needs a very good critical, analytical philosopher.
>
> Maxwell Bennett[151]

The compound assumption that human remembering must be undergirded by a process, or modular collection of processes, by which information that can be mapped to past experience is stored and retrieved, lies at the foundation of many attempts to study and model memory. Even those researchers who explicitly worry about these assumptions struggle to effectively purge them from their models, and many researchers realize that there is good reason to worry. Memory science has, again and again, demonstrated just how unlike the retrieval of archives the remembering process is.

It has long been recognized that memories were bound up with the process of remembering. Richard Semon coined the term "ecphory" to describe just this manifestation which was so intimately connected to "the memory" itself.[152] That is, the product of any instance of remembering depends profoundly on the context at the time of the remembering. Bartlett's

---

150 *Principles of Psychology*, p. 224.

151 *Neuroscience and Philosophy*, p. 174.

152 Semon 1925.

influential experiments showed that the remembering process is deeply constructive, and that retrieved details are surprisingly unreliable.[153] Bartlett and the constructive nature of memory fell out of favor for a few decades during the advent of the cognitivist revolution, but important work on construction and distortion continued again starting in the late sixties.[154] It was demonstrated that the retrieval of a memory can inhibit future retrieval,[155] and that it can also contribute to future retrieval.[156] It has been shown that the cues of a given retrieval can be incorporated into the memory itself,[157] and that what is remembered is thoroughly dependent on the context of retrieval—by virtue of both state-dependent memories[158] as well as misleading cues.[159] This dynamic relationship between memory trace and retrieval is difficult to reconcile with the storage metaphor. If an item is stored, but it is stored such that any accessing of it necessarily affects the item's constitution,[160] and it cannot be replaced unchanged after accessing it, we might well question the aptness of the storage model. This is what prompts Lynn Nadel to admit that few

---

[153] Bartlett 1932.

[154] Roediger 1996, p. 81. Bartlett's ideas have remained influential since.

[155] Brown 1968.

[156] Gardiner et al. 1973.

[157] McDaniel and Masson 1985.

[158] Godden and Baddeley 1975.

[159] The Loftus and Palmer experiments (1974) show that subjects who have seen a video of a car accident are systematically likelier to falsely recall seeing broken glass if asked about the cars "smashing" each other instead of "hitting" each other. Many later experiments confirm and elaborate on this kind of "postacquisition" interference.

[160] This is what impels Tulving to compare the engram to the Heisenberg uncertainty of an unobserved particle (Tulving 2007, p. 67).

memory researchers would privately defend the fixed memory trace paradigm, because "they know that memory is far more dynamic than our models have typically allowed."[161]

Amnesic patients demonstrate the ability to draw on past experience that they cannot consciously recall.[162] Even standard rememberers exhibit reconstruction and distortion in retrieval.[163] Any one "memory trace" can manifest in myriad ways,[164] and interference in various modes can destabilize even memories that had long been stable.[165] The stored item is now something that we cannot access without affecting, we cannot replace once accessed, can be present even when we do not deliberately access it, can be absent and distorted when we do deliberately access it, and can take very different forms depending on how it was accessed. It is no surprise that many memory scientists are concerned about the adequacy of the concept of storage.

The fluid, dynamic, and reconstructive nature of remembering, however, is well known—many of these experiments have been discussed, at length, for a half-century or more.[166] Many memory researchers have incorporated the constructive and labile nature of retrieval into the way they characterize memory and remembering. Even though data like these have given many researchers pause over their models, such considerations provide merely circumstantial evidence

---

161 Nadel 2007, 181. Following Nadel, I will sometimes adopt the talk of a "storage paradigm", although it must be recognized that these fall short of a "paradigm" in the Kuhnian sense.

162 See Toth 2000 for a review of these, some of which are parallel to the "blindsight" phenomena that have Interested philosophers working on perception.

163 Roediger and McDermott 2000.

164 Morris 2007, p. 31.

165 Sara 2000, p. 80.

166 "This new emphasis on neuroplasticity breaks radically with the idea of a permanent memory, which has piggybacked on the old '20th-century brain'" (Brockmeier 2010, p. 24).

against the deep-running computationalist assumptions above. An account of memory framed by storage and retrieval might easily include interdependencies and tendencies toward retrieval error (and most do). In fact, even the staunchest computationalist shouldn't expect human memory to share the clean modularity of digital computers—for a variety of reasons human minds and brains are simply messier than computers, and it would be surprising if this messiness did not result in various systematic errors. The best models of memory on offer have preserved the basics of the archive, but have also given due respect to the dynamics of memory as they have been revealed by experiment.

This is not to say that the lesson that memory is deeply fluid and dynamic is not one worth emphasizing. In fact, although this has been duly absorbed by those who study memory in science and philosophy, the outdated static and inert conception of memory is still frequently invoked by researchers *outside* the study of memory. When Clark and Chalmers write that the normal subject's remembering process is functionally analogous to searching through a notebook because the remembered belief "was sitting somewhere in memory, waiting to be accessed," they are just depending on a conception of memory that almost no contemporary memory scientist would support.[167] There are perhaps similar problems with the ways that philosophers characterize "stored beliefs" for epistemological purposes.[168] Philosophers are not alone in this, cognitive scientists and others who do not directly study memory, but who invoke it in their arguments, interpretations, and explanations tend to use a simple and outdated version of remembering that most specialists eschew. Of course, in one sense, this is not surprising: the development to a

---

167 Clark and Chalmers 1998, p. 11. This criticism is elaborated and well supported in Rupert 2004.

168 See Senor 2014 for a general argument that mechanisms of memory are relevant to arguments which draw on stored beliefs, or see Michaelian 2011, §3 for some specific implications for epistemology from a generative and constructive account of memory.

dynamic view of remembering has primarily taken place within the special studies of memory in the last few decades, and it is only natural that those within the multidisciplinary study of memory should have a readier and earlier grasp of this change than others. To this end, the more that philosophers and scientists that study memory are able to disseminate these revisions, the more quickly other researchers will be able to invoke more adequate characterizations of memory.

This is not the whole explanation, however. I will argue that even those who are proponents of this change, philosophers and otherwise, are struggling to understand the revised account of memory they are purveying. Some of the computationalist assumptions from chapter two run deep in our characterizations of mind and memory, and the "revised" view of memory that has resulted from this is something of a hybrid, and problematically so. As long as these newly developed accounts of memory are still so informed by computationalist frameworks, it can be no surprise that outsiders have a difficult time replacing the old computationalist models. In the following considerations I want to show, not only that this outside and simplistic conception of memory is belied by empirical findings—this I take to have been done several times already, even if the lesson has only seen mixed success—but also that the collection of revised accounts given by those who do study memory are insufficient to come to terms with these findings.

After studying the nature and extent of some of the results that have caused difficulties for an archival model of memory, and the attempts to incorporate them into more dynamic models of memory, I will argue that even those models that attempt to account for these ways in which the computational paradigm is not well suited to human memory fail to adequately retract important computationalist assumptions. Human rememberers are not messy computational systems that happen to tend toward systematic error: the "systematic errors" in question are part and parcel of the remembering process. Dynamicism and interdependency are not exceptions to the rule, these qualities of memory *are* the rule.

3.1 Disputes and impasses in memory science

Like all active scientific endeavors, the science of memory has its puzzles. And, like many other cases, much can be learned by the study of those puzzles that have resisted resolution in spite of sustained and variegated investigation. In one sense, of course, the primary goals of memory science themselves remain puzzling: many 21st century memory scientists admit that we still have only the faintest grasp on the basic nature and underlying physiology of memory.[169] There are also specific puzzles and disagreements, some of which have lingered without resolution for decades. Memory scientists have attempted to dissolve these conceptual standoffs, but often to no avail.[170] To better understand the reasons for this, I am going to present two long-standing, particular puzzles in memory science—both have been described as impasses, and both have induced researchers to try to rethink the conceptual frameworks within which these puzzles have arisen. The first is a debate about the nature of amnesia and other memory impairments. The second is a debate about what can and cannot be discovered in nonhuman animal remembering.

3.1.1 *Amnesia*

---

169 "Nevertheless, virtually nothing is known about the physiological processes underlying the act of remembering" (Sara 2000, p. 76); "We have only a vague idea about memory as a whole. We cannot even say if there is such a thing as memory (or a memory, or specific memory systems) at all (Brockmeier 2010, p. 5).

170 Frustration toward these back-and-forth conceptual issues drives some researchers to hopelessness concerning the language of the discipline. Michael Hasselmo writes, "The discussion of memory concepts is useful, but ultimately verbal terms are insufficient for describing memory function" (Hasselmo 2007, p. 123).

The study of memory impairment is beset by singular difficulties. The last several decades of research have yielded significant advances in our ability to induce, to identify the causes of, and to predict the effects of, various amnesias. Yet memory scientists have been debating the same basic question about the nature of amnesia for the duration: whether these impairments reflect a failure to store what is needed, or a failure to retrieve what is stored.[171] Despite four decades of lively and sustained debate on the subject, researchers continue to be divided between the "storage view" and the "retrieval view" of amnesia. This is an especially awkward state of affairs given the continuing developments in amnesia induction and prediction: what is taken to be the basic nature of the underlying process remains unresolved despite our increasingly rich grasp of the characteristics of impairment phenomena. This stubborn inability to distinguish storage failure from retrieval failure has led several researchers to worry whether the distinction has been adequately framed,[172] and it has led some to doubt altogether the viability of explaining memory deficits in terms of storage and retrieval.[173] After decades of failed attempts to empirically separate storage failure from retrieval failure, it is unsurprising that we might question whether these are even *conceptually* distinct—whether the notion of a distinct, stored memory trace is a coherent notion at all, and whether positing this is needed to explain memory and impairment.

Challenging the coherence of, and need for, conceptually distinct notions of storage and retrieval in accounts of memory is not a new enterprise. Especially among researchers influenced by Wittgenstein, and perhaps increasingly in recent years, a number of criticisms have been leveled

---

[171] See, for example, the entirety of the 2006 issue of *Learning & Memory* (Volume 13, number 5) for a thorough survey of this continuing debate, which has foundations in the science of memory that run at least as far back as Weiskrantz 1966.

[172] E.g., Nader and Einarsson 2010, p. 33.

[173] E.g., Squire 2006; Matzel and Miller 2009.

against the expectation that a proper account of memory will be in terms of an underlying process of storage and retrieval.[174] Wittgenstein himself remarked on the problems inherent in storage and retrieval as explanatory of memory phenomena, decades prior to the extended debate over amnesia. We will see below that many of these remarks offer powerful criticisms. Some of these derive from Wittgenstein's overarching critiques of the impulse to posit hidden mental processes in general, and others are explicitly and particularly concerned with the vacuity of positing stored representations in explaining memory. Wittgenstein claimed that carving the process of remembering—the process with which we are already familiar—into some further, hidden, complex set of inner processes is a misguided endeavor.[175] The remembering process, according to Wittgenstein, is just the one we already know, and would be a mistake to seek to delineate separate, hidden components to this process. Scientists currently working on memory impairment are beginning to agree, as Larry Squire demonstrates in his recent article, "*Lost forever or temporarily misplaced? The long debate about the nature of memory impairment*":

> "...the terms "storage" and "retrieval" are not particularly apt for describing recovery from retrograde amnesia. If those terms are to be used, one would better describe the deficit as a transient loss of the mechanism that supports both storage and retrieval."[176]

I will contend that a Wittgensteinian dissolution of hidden mental processes is just what the theoretical gridlock in the amnesia debate requires. This is intimated by the fact that various researchers are slowly converging on positions reminiscent of Wittgenstein's, but will be best

174 Randall 2007; Hardt et al. 2009; Brockmeier 2010; Sutton 2014.

175 "It is only that 'There has just taken place in me the mental process of remembering…" means nothing more than "I have just remembered…' To deny the mental process would mean to deny the remembering…" (Philosophical Investigations §306). These remarks will be returned to below.

176 Squire 2006, p. 5.

demonstrated by bringing these Wittgensteinian considerations to bear on a recent and innovative attempt to empirically resolve the debate about storage and retrieval.

Empirical work designed to study memory impairment faces immediate difficulties. Models of amnesia often target the kinds of memory dysfunction present in cases like the famous amnesic H.M., who suffered dense anterograde amnesia that affected his ability to form new episodic memories, or the retrograde amnesias exhibited by Alzheimer's patients' difficulties with autobiographical remembering. These are largely impairments of *declarative* memory—abilities to verbally recall past episodes or learned facts.[177] However, much of the laboratory work that informs these models and characterizations uses non-human animal subjects, which precludes any direct study of declarative remembering.[178] The study of memory impairment in animal subjects investigates *nondeclarative* memory systems and their failures, and researchers then use these data to inform their models of amnesias that extend to declarative cases. This practice probably involves problematic inferences that deserve further scrutiny. In many instances experimenters technically draw conclusions concerning only the variety of impairment present in the study, but use language that is all too easily appropriated for generalization to human amnesics.[179] Even wholly nondeclarative induced amnesia can be difficult to generalize—particular contexts that

---

[177]In fact, H.M.'s *competencies* in non-declarative remembering tasks provided some of the original evidence for the separation of different memory systems.

[178]There are controversial claims of certain non-human animals demonstrating "episodic-like" memory capabilities, but even if they are rightly so called, these are generally not the capacities studied in impairment research.

[179] Hardt et al. address this concern but still use the term 'amnesia': "'Amnesia' does not refer to a comprehensive impairment affecting all aspects of the memory distributed across the brain. The question is whether storage or retrieval is impaired in the targeted memory system. Thus, amnesia, as used in the studies reported here, refers exclusively to dHPC-mediated representations and processes involved in contextual fear conditioning" (Hardt et al. 2009, 225). We will return to this point in the next section.

produce impairment behaviors are often difficult to replicate in other, even similar, contexts.[180] Generalized claims about the nature of amnesia may be contentious, but even if the question of storage or retrieval failure could be resolved for just one type of memory impairment in one set of conditions, this would represent progress in our understanding of memory impairment.

In 2009, Oliver Hardt and colleagues gathered experimental results that they claimed to decisively support the "storage view" in at least one type of induced amnesia.[181] The authors of the study are sensitive to the erstwhile difficulties in disentangling the storage view from the retrieval view (the article is even titled *Storage or retrieval deficit: The yin and yang of amnesia*):

> Two orthogonal interpretations exist for memory recovery after amnesia induction, and the currently available evidence does not allow deciding between them. ...Both positions can explain most outcomes of experiments that use memory recovery as a criterion for the presence or absence of memory after amnesia induction.[182]

Nonetheless, Hardt and colleagues claim to have divided the yin from the yang with a new experimental paradigm, and this methodological advance is emphasized as the more important conclusion to be gleaned from the experiment. This study provides an example well suited to Wittgensteinian critique. This is not because Hardt et al. make glaring conceptual errors or sweeping theoretical claims—in fact, the article demonstrates a good deal of subtlety and sensitivity to conceptual and theoretical issues. The experimental design is innovative, but ultimately falls prey to the very concerns the authors were striving to avoid. If Wittgenstein is right, such will be the fate of any attempt get between the surface phenomena of remembering and its supposed underlying, component mental processes.

---

180 Dudai 2004, pp. 70-71.

181 Hardt et al. 2009.

182 *Ibid*., p. 224.

One basic problem for memory scientists seeking to resolve the amnesia debate is that what are taken to be competing hypotheses seem reluctant to yield distinct empirical predictions. When the amnesic subject fails to remember in any given context, it might always be that the stored trace remains intact but inaccessible. When the amnesic subject succeeds in remembering, this can be taken to signify a retrieval difficulty that has been overcome, but it can also be interpreted as a consequence of new learning in addition to whatever incomplete trace was stored. Amnesia is never *complete* amnesia—there are always some residual effects from an experience even in cases of the most profound memory impairments.[183] This means that neither the presence nor the absence of memory recovery succeeding impairment is inconsistent with either of the deficit hypotheses:

> The issue of whether experimental amnesia is a storage or retrieval deficit has been debated for decades without resolution, primarily because the paradigms that have been used to probe for whether amnesia is a storage or retrieval deficit are unable to differentiate between these two interpretations.[184]

As various researchers have noted, the overarching problem is that memory impairment is a behavior-level phenomenon, while the component processes in question are taken to be cognitively internal.[185] At the behavioral level, researchers are ultimately restricted to whether remembering has failed or succeeded in specific instances. While particular combinations of successful and impaired remembering abilities provide criteria for characterizations of *types* of remembering, there is no behavioral demonstration that a memory has been stored other than just a successful act of remembering. Memory storage and retrieval are not supposed to be ways in which we remember

---

[183] Gold and King 1974. Schacter 2008, p. 160.

[184] Nader and Einarsson 2010, p. 33.

[185] Squire 2006, p.3.

or fail to remember, but rather underlying process components that are in some sense responsible for the behavioral phenomena of memory.

It should be noted that these infra-behavioral subprocesses are not *neural* processes either. That is, it is not the case that two different biological processes have been observed, and researchers are trying to correlate the presence or absence of these biological processes with various behavioral phenomena. Just how storage and retrieval are neurally implemented is taken to be an open question, and a controversial one in its own right.[186] Fitting these psychological subprocesses in to what neuroscientists have observed at the neural level is proving to be just as problematic as fitting them in to what psychologists have observed at the behavioral level.

Just following the above remarks on the denial of the mental processes of remembering, Wittgenstein goes on in the *Philosophical Investigations* to explicitly question just this move:

> How does the philosophical problem about mental processes and states and about behaviourism arise?--The first step is one that altogether escapes notice. We talk of processes and states, and leave their nature undecided. Sometime perhaps we'll know more about them—we think. But that's just what commits us to a particular way of looking at the matter. For we have a certain conception of what it means to learn to know a process better. (The decisive moment in the conjuring trick has been made, and it was the very one that seemed to us quite innocent.)--And now the analogy which was to make us understand our thoughts falls to pieces. So we have to deny the yet uncomprehended process in the yet unexplored medium. And now it looks as if we had denied mental processes. And naturally we don't want to deny them.[187]

The storage and retrieval processes that are said to underlie remembering and forgetting are exactly so: we have talked of these processes and states, but we have left their natures undecided. We have a conception of what it might mean to come to know these processes better, but in actually attempting to parse these inner processes we seem to only encounter impasse. The

---

186 "One of the major aims of modern memory research is to locate the physical substrate of memory (also referred to as 'memory trace' or 'neural substrates of memory') in the brain" (Sakaguchi and Hayachi 2012, p. 1).

187 *Philosophical Investigations* §308. *Cf.* James 1890/2007, p. 224.

analogy of an archive or storage system has only made it more difficult to understand what amnesia really is, and the interpretation and re-interpretation of relevant experiments appears unending.

In their recent study, Hardt and colleagues sketch this experimental impasse and proceed to advocate a "paradigm" that cuts through these difficulties:

> To resolve this stalemate, a paradigm is needed that (1) allows the storage impairment to positively demonstrate the absence of memory, which is not equal to the absence of recovery, and (2) equips retrieval and storage views of amnesia with opposing predictions, making empirical disambiguation possible. We here describe a method that meets both requirements.[188]

The innovation presented by Hardt et al. takes advantage of a correlation between the neurobiological and behavioral characteristics of learning in rats. Certain receptors (NMDAr) in the dorsal hippocampus are necessary for successful fear conditioning in first-time learning, but they are not necessary in second-time fear conditioning in a new context. In rats that have already been fear-conditioned in one context, this receptor is not necessary for successful conditioning in a second context. The presence of an inhibitor that blocks these receptors—AP5[189]—usually prohibits first-time learning but not second-time learning. So the proposal is to induce amnesia in the first instance of learning (with a protein synthesis inhibitor—ANI[190]—known previously to block fear-conditioned learning in rats), then to see if blocking the NMDAr antagonists prohibits learning in the second instance. If blocking these receptors also prohibits second-time learning after inducing amnesia the first time around, argue Hardt et al., the stored trace must be absent, and so the induced amnesia is a failure of storage. If blocking these receptors does not prohibit second-time learning, then the stored trace must be intact and the amnesia induced by the protein

---

188 *Op. Cit*., p. 224.

189 2-amino-5-phosphonopentanoic acid.

190 Anisomycin.

synthesis inhibitor is a failure of retrieval. Granted, this could only decide the nature of the impaired fear conditioning induced in rats by a particular protein synthesis inhibitor, but hitherto *no* variety of amnesia or impairment has been shown to be a failure of storage or retrieval.

The results of the study, according to the above interpretation, support storage failure. That is, in the experiment Hardt and colleagues ran, the NMDAr antagonists did prohibit second-time learning after the amnestic agent was administered the first time round. For the amnesic rats, the second instance of learning was neurobiologically like a first instance. Therefore, according to the authors, the stored memory trace was not intact after the induced amnesia. If it had been intact, and the induced amnesia had been a matter of retrieval failure, then the second-round learning would not have been affected by blocking the NMDA receptors. According to the amnesic rat's hippocampus, the second instance of learning *just was* the first instance, and so there was no stored trace retained from the first instance.

The problem here is that it was already understood that there was *some* sense in which the second-time learning was first-time learning—this is simply what amnesia amounts to. To claim that a vulnerability to NMDAr antagonists in the second learning instance implies the absence of a existing stored trace is to claim that the antagonists would not inhibit retrieval if only there existed the stored trace to be retrieved. The retrieval side of this vulnerability is taken to be immediate, as though any stored trace would be automatically retrieved to the point that the NMDA receptors were no longer necessary. We might have cast the whole experiment in the reverse by taking storage in such an automatic fashion: if what was stored could just be accessed, the opponent might argue, then AP5 would not block learning the second time around. The remembering failure in the second instance can support either interpretation.

Hardt et al. do briefly express concern about the possibility of this kind of alternate interpretation, but they do not adequately engage with it.[191] In a commentary on the Hardt et al. paper, Louis Matzel and Ralph Miller point out the difficulty:

> If the vulnerability of the second training event to AP5 depends on the retrievability of the original memory, then Hardt et al.'s observations would be entirely explicable in terms of ANI having induced a retrieval failure. Hardt et al.'s preferred interpretation of these effects presupposes that the basis for ANI-induced amnesia is known, when in fact these very experiments were intended to determine the basis of this effect. These complications are precisely analogous to the one asserted by Hardt et al. to have plagued previous efforts to distinguish storage from retrieval interpretations of experimentally induced amnesia.[192]

While there is genuine innovation in attempting to distinguish components of the cognitive process of remembering by tracking neurobiological signatures that have been correlated with successful and unsuccessful remembering, these could only demonstrate the success or failure of *subprocesses* like storage and retrieval *if such neurobiological phenomena were already correlated to these subprocesses*. NMDAr antagonists and protein synthesis inhibitors have been demonstrated to impede remembering because these have been correlated with behavioral memory failure, the criteria for which are well known. In order to track the chemical or neurobiological signatures of storage or retrieval, these would have to be correlated with criteria for *subprocess* success and failure. But this, of course, is just the problem: we do not know what it would look like for storage or retrieval to fail, other than simply what it already looks like when memory fails.

This suggests two lines of criticism. First, if these infra-behavioral, but supra-neural, subprocesses are not brought into the attempted explanation because they are familiar to us from

---

191 "Post hoc interpretations of the data can describe the extinction effect as memory erasure and the behavioral impairments induced by AP5 and ANI as retrieval impairments. However, these post hoc interpretations do not supply any novel predictions. Therefore, they remain philosophical positions that cannot be empirically addressed" (Hardt et al. 2009, p. 229). Just what it is that makes these "philosophical positions" post hoc, in any way that the Hardt et al. interpretation is not so, is not further clarified.

192 Matzel and Miller 2009, p. 671.

the phenomena of remembering and forgetting, and they are not brought into this explanation from brain science, what are they doing here? If carving remembering and forgetting into the underlying mental processes of storage and retrieval causes such difficulty, can we do without? Must we posit the intact stored trace in order to explain remembering and failing to remember? Rather, was this not just "the decisive moment in the conjuring trick" of which Wittgenstein warns? The remark from Larry Squire that was mentioned earlier proceeds in this direction, as he advocates a view of amnesia as the "loss of the mechanism that supports *both storage and retrieval*."[193] This does away with the notion that these components of the remembering process are needed at all in our explanation of amnesia.

Second, in addition to asking about whether these component processes are *needed*, we may also be impelled to ask whether they are *viable*. The conceptual separation of storage from retrieval seems persistently problematic, and the two can seem inextricable not only empirically, but conceptually. In the very same article, Hardt et al. touch on this worry as they explain the amnesia impasse:

> The retrieval impairment account of experimental amnesia, at the most fundamental level of analysis, needs to assume basic storage processes as well, as the information required to retrieve memories, for example, the retrieval "cues" themselves, must somehow be integrated into the networks that mediate retrieval. In other words, retrieval requires the storage of information that permits retrieval.[194]

The retrieval component requires storage; a "retrieval" system with nothing stored could not get off the ground. If a retrieval subprocess must actually include both retrieval and storage, it is far from clear that the *conceptual* isolation of storage from retrieval is feasible.

---

[193] Squire 2006, p. 5, emphasis added.

[194] Hardt et al. 2009, p. 228.

Finding fault with the need for, and the viability of, the inner mental processes that are supposed to underlie and explain the surface phenomena of cognition—and in particular, memory—has a long tradition in the work of Wittgenstein and those influenced by Wittgenstein. These criticisms that memory scientists have been recently offering are converging on threads running through 20th-century philosophy of psychology, and a study of the powerful considerations Wittgenstein provides helps to sharpen the very arguments on offer in present-day memory science.

3.1.2 *Language use and the character of episodic memory*

The second impasse in memory science that deserves our attention is one that concerns episodic memory in particular. Episodic memory is taken to be a form of declarative remembering—that is, its retrieval takes the form (or at least *can* take the form of) utterances. Episodic memory is also taken to be one of the memory systems, a particular conjunction of neural and inner cognitive processes that allows us to recall episodes from our past. These two facts taken together, however, seem to imply that we can either know ahead of time what we will and will not find in the brains and minds of nonlinguistic animals, or that we have posited a process that can be neither verified nor falsified in the case of nonlinguistic animals. This is especially problematic given that the learning and remembering behavior of some intelligent, nonhuman animals is similar in many regards to the behavior of humans, and our present concepts of memory leave us unable to characterize these findings. Whether nonhuman animals should be said to *episodically* remember has been a pivot point around which new focus on memory concepts has recently revolved and counter-revolved, and so it benefits a study of memory and its components to turn to the comparative psychology of remembering episodes.

Scrub jays, like most members of the family *Corvidae*, exhibit remarkable intelligence. One capacity that demonstrates this is the jays' ability to distinguish their own food caches on the basis of their contents, locations, and relevant time intervals.[195] This last, especially, has led cognitive scientists to claim that the jays are able to recall the 'what, when and where' of specific episodes,[196] an ability that hitherto had been claimed to be exclusively human.[197] In response, critics within the cognitive sciences have responded by offering alternative explanatory models of the jays' behaviors,[198] and have defended the unique status of human rememberers.[199] Similar experiments have shown sophisticated memory capacities in rhesus monkeys,[200] provoking similar controversy.[201] Most recently, the same discussion has been of the experimentally confirmed capacities of chickadees.[202] The debates continue.

Episodic memory—that is, the family of those abilities that consciously invoke past episodes—is an immediately problematic subject of inquiry in nonlinguistic subjects. Endel Tulving, who is credited as the source of this classificatory concept and whose definition is still the most widely employed, explicitly characterizes episodic memory in terms of *consciousness*. For Tulving, the fact that one is subjectively aware that a remembered episode is personal is what

---

[195] Clayton and Dickinson 1998.

[196] Griffiths et al. 1999.

[197] Bischof-Kohler 1985; Tulving 1995.

[198] McCormack 2001.

[199] Tulving 2005.

[200] Hampton 2001.

[201] Suddendorf and Busby 2003.

[202] Feeney et al. 2009.

separates episodic from mere *semantic* memory—that is, the ability to remember facts or true statements.[203] This distinction seems credible enough; William James counted as memory only those phenomena of which we were so conscious,[204] and introspection alone at least appears to substantially divide remembering, for example, one's seventh birthday party and remembering *that one had* a seventh birthday party. The problem is that the "autonoetic" consciousness that Tulving makes a criterion of episodic remembering seems to be accessible only by introspection or linguistic query, at best. Even if episodic memory were a unitary, definite, inner process that we humans did in fact share with other species—and each of these is a problematic assumption in its own right—it appears that we would forever lack the capacity to confirm this, unable to interview our nonlinguistic counterparts. It was the impossibility of meeting this consciousness criterion that has prompted Nicola Clayton and Anthony Dickinson, in describing the results of their well-known experiments with scrub jays, to characterize the memory abilities of the jays as 'episodic-like' rather than simply as episodic memory.[205] Furthermore, it prompts some researchers to despair of the program altogether, since the consciousness criterion "presents an insurmountable barrier to demonstrating this form of memory in animals."[206]

Clayton and Dickinson's scrub jay experiments make a good start—the jays reliably choose to recover their caches of worms rather than their caches of (considerably less perishable) peanuts only when it has not been long enough for the worms to have rotted. That is, without any additional input about the state of the peanuts or worms, the jays will choose the worms at all and only those

[203] Tulving 1983.

[204] James 1890/2007, p. 648.

[205] Clayton and Dickinson 1998.

[206] Griffiths et al. 1999, p. 77.

times when they have not been buried long enough to have spoiled. Other birds lack this ability, and must actually detect the state of the worms in the present moment in order to forego them for the non-perished peanuts.[207] It is indeed tempting to explain the jays' behavior by simply invoking their ability to consciously recall the event of burying the worms, and to decide their course of action based on the results of this successful recollection. In fact, we would not hesitate to impute such a process to a human that we witnessed behaving similarly. If my roommate were to come upon a sandwich that he had left on the counter, and I saw him consider the sandwich, look contemplative for a moment, and then decide to eat the sandwich, I would very reasonably conclude that he had brought to mind the past event of making the sandwich and had established that this event was not so long ago that the sandwich was no longer worth eating.

The problem, as critics have pointed out,[208] is that any sensitivity to the duration itself will suffice to explain the changes in behavior—the jays need not *recall* past episodes, they only need be affected in some way by the event and duration. Roommates can know whether it has been more than an hour that a sandwich has been sitting on the counter, but sleep timers can also "know" whether the radio has been playing for more than an hour, and nobody is tempted to impute conscious recall to the sleep timer. This difference is stressed by Morris and Frey, albeit with a certain optimism:

> To model episodic memory in animals more effectively, the challenge is therefore to devise new tasks in which having a `recollective experience' would be helpful (or even required). More formally, such a task should distinguish between changes in behaviour that occur because an animal remembers some prior event and changes in behaviour that merely happen because some prior event has occurred. This distinction, although subtle, is absolutely fundamental to the claim that animals possess `elements of episodic memory'.[209]

---

[207] In fact, the jays' sensitivity to duration seems more finely tuned than comparable unaided human sensitivity (Wearden et al. 1997).

[208] See Suddendorf & Corballis 1997. A similar response is canvassed in McCormack 2001.

[209] Morris and Frey 1997, p. 1495.

The reason such a task may be required is that comparative psychologists are beholden—at least to some extent—to the methodological precept known as Morgan's Canon. Named after the pioneering comparative psychologist C. Lloyd Morgan, this principle urges the researcher never to invoke a 'higher level' psychological explanation when a 'lower level' one will do.[210] That is, if the behavior of the jays can be explained without invoking high-level cognitive capacities like episodic recall, it should be so explained. The difficulty is this: absent linguistic confirmation of episodic recall, it is not clear what sort of evidence *would* require the higher level explanation. This is not just a consequence of the consciousness criterion built into the definition of episodic memory: even without that criterion, critics claim that we are not in these cases required to posit anything like episodic recall.[211] Episodic memory is taken to constitute its own memory process, and that process is supposed to be something separate from, underlying, the mental images or verbal stories *produced* by this process. This makes it seem as though nonlinguistic animals would be useful subjects in studying this process—but instead, the comparative psychologist's episodic remembering seems to be an inner process that *no experiment could ever detect*. This is a sure sign that there is good reason to be suspicious of the concepts involved.

Researchers are divided on what to do with the apparent similarities between what is called episodic remembering in humans and the duration-sensitive behaviors demonstrated among nonlinguistic animals in experiments like those of Clayton and Dickinson. It may, following

---

[210] Morgan 1903. Morgan's own endorsement of the canon, the precise meaning of higher and lower levels, and whether the canon is a desideratum at all are each disputed claims. (Cf. Hans-Johann Glock 2013, p. 164: "Although it is a well-known methodological principle of comparative psychology, it is far from obvious what Morgan's canon actually entreats us to do, and on what grounds.") Nonetheless the basic notion of the canon is intuitively acceptable. Notably, Morgan's work in comparative psychology is often seen as a precursor to behaviorist methods in general psychology.

[211] Suddendorf and Corballis 1997, p. 391.

Morris and Frey's sanguine attitude, turn out to be the case that an experiment that is clever enough will require a genuinely episodic explanation, or it may be that the search for such an experiment is a fool's errand. To examine just what it is we seek in testing for nonlinguistic thought, let us turn briefly to the much-discussed and problematic relationship between language and thought.

As William James remarked over a century ago, "The question whether thought is possible without language has been a favorite topic of discussion among philosophers."[212] The thesis that thought depends on language takes many forms, but one starting point is the work of Donald Davidson. In *Thought and Talk*, Davidson claims that neither thought nor language enjoys conceptual priority over the other, and he also explicitly claims that only creatures with language can have thoughts.[213] Peter Carruthers responds to this claim by pointing out that Davidsonian conceptions of the mind are so at odds with the basic assumptions of cognitive science as to not warrant consideration:

> But these arguments all depend, in one way or another, upon an anti-realist conception of the mind—claiming, for instance, that since we cannot interpret anyone as entertaining any given fine-grained thought in the absence of linguistic behavior, such thoughts cannot even exist in the absence of such behavior. Since the view adopted in this paper—and shared by most cognitive psychologists—is quite strongly realist about thought, I do not propose to devote any time to such arguments.[214]

Carruthers is surely right that any view of thought as consisting in interpreted speech behavior necessarily precludes those creatures who do not speak. He may also be correct to ultimately construe Davidson's larger approach to propositional attitudes as such a view. However, in those writings where Davidson is most explicitly defending his rejection of

---

[212] James 1890/2007, p. 266.

[213] Davidson 1975, pp. 156-157.

[214] Carruthers 2002, p. 661.

nonlinguistic thought—*Thought and Talk* and *Rational Animals*[215]—the arguments he presents do not depend on the inability to interpret the nonlinguistic creature, but rather on the nonlinguistic creature's inability to interpret others. Davidson characterizes the "chief thesis" of *Thought and Talk* as the claim that "a creature cannot have thoughts unless it is an interpreter of the speech of another,"[216] and he is explicit in entertaining the notion of a thinking being who is himself speechless.[217] Only through interpretation, argues Davidson, can we acquire the necessary framework of truth and error—without these, we cannot have the concept of belief. Furthermore, without the concept of belief, we cannot entertain beliefs, which are necessary for thought.

So Carruthers is too quickly dismissive: even if we reject Davidson's antirealism about the mental, his argument for the necessity of error for belief and thought appears to stand independently of any such commitments. It is not merely because we lack the fine-grained linguistic expressions to ascribe beliefs to a speechless creature, but rather because the creature itself has not undergone what is required in order to form beliefs. Indeed, when Davidson speaks in terms of the concepts had by a creature and the cognitive abilities thereby gained or absent, he sounds just as realist about the mental—at least for the moment—as Carruthers would like. Carruthers himself endorses a certain dependency of thought on language: his own view is that the mind is modular, and that cross-module thought requires the generativity brought about by language acquisition.[218] This move is helpful in one respect: it specifies the notion of thought that we are attempting to characterize.

---

215 Davidson 1982.

216 Davidson1975, p. 10.

217 *Ibid*., p. 23.

218 Carruthers 2002, p. 665.

Part of the difficulty in teasing apart varying theses concerning the relationship between thought and talk is simply that 'thought' is an overwhelmingly broad and variegated concept. If ever there was a candidate for a family resemblance concept, 'thought' is surely that candidate. A flat-out 'yes' or 'no' answer to the question of whether non-human animals can think seems trivial, ridiculous, or both at once: there is certainly some sense in which a scrub jay and a human are of a single cognitive kind (as opposed to, say, a rock) and there is certainly some sense in which a they are of differing kinds. Rather than contend over the most natural use of the term, or risk talking past one another, it must be preferable to be more precise in the specific cognitive capacities that we are investigating. To this end, a focus on episodic memory will be helpful: it is a cognitive capacity that is relatively well-defined and well-recognized, and it happens to be one that already enjoys a contested status as to whether it belongs solely to humans. In this way, an examination of what it means for nonlinguistic creatures to episodically remember will not only aid those scientists studying memory, but also those philosophers studying thought and language. If Davidson is right, and belief requires the interpretation of language, it may be that we can know ahead of time that nonlinguistic animals do not recall episodes (if such recall requires belief), or we may devise an experiment that demonstrates that they do so recall, thereby undermining Davidson's reasoning. If Carruthers is right, and cross-module, non-domain-specific thought requires language acquisition, this may disallow nonlinguistic creatures from exhibiting episodic memory, or we may likewise discover this capacity and work backward to countermand Carruthers' view (or the modularity of episodic remembering). However, it is also important to keep in mind that the question whether nonlinguistic creatures episodically remember may admit of neither an affirmative or a negative answer—Peter Hacker offers a Wittgensteinian reason to pause before searching for the truth or falsehood of statements like this:

> The limits of thought are the limits of the behavioural expression of thought. A creature can intelligibly be said (truly *or falsely*) to think that things are so just to the extent that its behavioural repertoire includes such forms of action and response that would warrant the ascription to it of *thinking things to be so*. The limits of what a non-language-using animal can be said to think are the limits of its cognitive expressive behavior.[219]

That is, whether the question of the Jays' having episodic memory can even meaningfully be answered may depend on the precise conceptual details of what we mean by 'episodic memory.' If to "episodically remember" is just to engage in behaviors that are unavailable to the scrub jays—such as recounting past episodes verbally—then it will be very easy to restrict nonlinguistic animals from episodic remembering: we have strong evidence that nonhuman animals cannot tell stories about anything, much less about their pasts. If to 'episodically remember' is to engage in certain cognitive or neural processes that sometimes happen to *result* in language use, then we may be able to directly test for those neural or cognitive processes. If we do not know whether episodic memory is the former or the latter, then we will be condemned to equivocation. The assumption that remembering is an inner cognitive process encourages researchers to seek evidence of this inner process even among nonlinguistic creatures, but formulating this hypothetical process with precision sufficient for empirical consequences remains troublesome.

It is worth noting, in passing, that the details of this investigation may hold implications for the relationship between thought and language. Imagine that the search for an inner cognitive (or neurobiological) process which is or supports episodic memory is not misguided, so that we can go about searching for episodic remembering among nonlinguistic creatures. This might first require some conceptual cleanup on our part: our notion of episodic memory is still in development, and confused concepts will make for confused experimentation. Nevertheless, the optimistic researcher can claim that, armed with a refined concept and savvy experimental design,

---

219 Hacker 2013, p. 15, original emphasis.

we will be able to test for episodic memory without requiring language use. One shortcut to this result would be via cognitive neuroscience: those researchers who believe that identifying the neural correlates of episodic memory is just a matter of time will no doubt suspect that comparative neuroscience might circumvent all of these troublesome language limitations.

And what might we find if we were able to perform any of these experiments? Here there are just two basic alternatives, but each presents further possible implications. First, we might find that there are no non-language-using creatures who episodically remember. Davidson's, Carruthers' or another argument for the language-dependence of thought might be right, and episodic memory could turn out to depend on a sort of thought that requires language. It is worth noting that the discovered dependence in this case will be non-conceptual. That is, if language were conceptually necessary for episodic memory, then the connection would not be discovered experimentally. Rather, if experiments demonstrated that there were no nonlinguistic episodic rememberers, though there could have been, this could only—at the strongest—demonstrate some kind of non-conceptual necessity. Language may just be nomologically necessary for the development of episodic memory.[220] There is at least circumstantial evidence for this view from developmental psychology—it is well known that at a certain developmental stage, children appear to have facility only with the simple past tense, and reliably demonstrate that they lack real comprehension of even this simplified concept of the past.[221] The acquisition of the linguistic mastery of the past tense is roughly contemporaneous with the age—around 3-4 years old—that marks the beginning of accessible episodic memory. This is consonant with the language dependence of episodic memory capacities, but of course it provides similar evidence for a

---

220 Clark ascribes this sort of nomological dependence on "linguaform resources" to higher order thoughts, among other complex cognitive capacities (Clark 2001, p. 144-145).

221 McCormack 2001, p. 190.

dependence that runs the other direction: some researchers have argued that it is the developmentally acquired ability for 'mental time travel' that provides the scaffolding for language acquisition.[222] Indeed, it would be philosophical considerations like those of Davidson or Carruthers that would decide the direction of this dependence.

We may find, upon refinement of our concept and successful experimental design that there is a robust sense of episodic memory that we share with other species. Again, deciding whether episodic memory is contingent on belief will make or break Davidson's claim: given the inclusion of a criterion like Tulving's autonoetic consciousness, which specifically denotes self-knowledge, a dependence on belief will follow. If this holds, and episodic memory is found in nonlinguistic beings, Davidson's account is flawed. Carruthers' weaker claim, that cross-domain thought requires language, is more difficult to obtain straightforward experimental predictions from. A collection of domains such as self-knowing, past-knowing, and the practical details of the remembered event appears 'cross-domain' in some sense, but if Carruthers' claim is to do predictive work, this sense will have to be specified. Part of the work of cleaning this up has been done, dialectically, by José Bermúdez, who has argued that modular and cross-modular thought is available to creatures even despite the absence of linguistic tools.[223] Although it is still conceptually premature, the status of modularity of thought and its relationship to language will directly inform models of comparative psychology and the cognitive capacities of nonspeaking creatures. More work on whether episodic memory displays this kind of cross-modularity is needed.[224]

---

[222] Suddendorf and Corballis 1997, p. 133.

[223] Bermúdez 2003.

[224] Carruthers acknowledges this, but his present examples of cross-modularity are comparatively anecdotal.

But either of these would be truly remarkable. Finding that there was an inner process of episodic remembering, which these jays lack or demonstrate, would mean that some correlation had been established between the inner process and the observable behavior of the organisms or their brains. How could this have been established? Part of the reason that this supposition can seem plausible in any given context of investigation is that the act that establishes this correlation can be conveniently pushed to whichever side of the process is opposite the investigator. That is, the cognitive neuroscientist can seek the underlying process of "episodic remembering" in the brain, on the supposition that the inner *cognitive* process has been successfully correlated to behavior by the psychologist. Conversely, the psychologist can work toward correlating the behavior with "inner processes" on the supposition that the "inner process" is, or is correlated with, an identifiable neural process which the jays either have or lack. The difficulties researchers have in understanding even what would count as episodic remembering in nonlinguistic creatures is just a way to reveal the often disguised fact that neuroscientists and psychologists are each assuming that the correlation has been established at another level. Once again, the assumed hidden process of remembering resists adequate empirical characterization. Memory scientists re at a loss as to how to design or interpret experiments that conclusively demonstrate the character, or even existence, of the inner processes of memory.

3.2 Revisions to the standard models

The above difficulties are well known, and they have generated a collection of responses and revisions. Researchers across many of the disciplines that take an interest in memory have doubted the adequacy of the traditional archival conception of memory for decades, and have responded to these worries in various ways. Daniel Schacter dismisses memory as a "hypothetical store in which information is held" as an implausible core concept of memory, citing the broad disfavor this hypothesis enjoys among memory scientists.

> The notion of a 'memory store', though widely used in everyday parlance about memory, is not a concept to which many contemporary students would subscribe. Cognitive psychologists have long abandoned the idea of a 'memory store' in favor of a more processing-oriented view of memory, and most neuroscientists conceive of memory storage as a distributed process, with a variety of brain regions participating in storage.[225]

However, marks of the 'memory store' Schacter is attempting to reject can be noted in his own characterization of its replacement and throughout Schacter's accounts of memory.[226] Lynn Nadel claims that the notion of a fixed memory trace "stands as a barrier to further progress in understanding the nature and underlying bases of memory."[227] Nadel goes on, in the same article, to characterize memory in terms of a database of encoded representations.[228] This pattern is common in memory science: many researchers reject notions of storage in name, but then fall back on its conceptual trappings as an account is engaged. As Yadin Dudai admits, "Scientific paradigms might be more static and resistant to change than brains."[229] Some of the resulting admissions, caveats, and revisions have been clearer and more helpful than others.[230]

---

225 Schacter 2007, p. 25.

226 For example, Schacter also wrote, one year later that the seemingly innocent idea that memories exist in a vast storehouse is "fundamentally misleading." He then adds, in the *following sentence*, "Our experience of remembering an event does, of course, partly depend on information about the event that has been stored in our brains" (Schacter 2008, p. 16).

227 Nadel 2007, p. 179.

228 *Ibid*., p. 180.

229 Dudai 2004, p. 79.

230 Although it is motivated by some of the above concerns, I find Norman Spear's, apparently influential, definition of memory as a "multidimensional conglomerate of attributes" bewildering and unhelpful (Spear 1978, p. 56). For example, Susan Sara's otherwise clear article on the neurobiology of remembering is only impeded by borrowing Spear's phrase to explain that the physiological state of the organism is "incorporated into the conglomerate of memory attributes and exerts control over retrieval of the memory" Sara 2000, p. 75).

In many cases, these considerations have merely induced researchers to slightly alter their language, as psychologist Jens Brockmeier points out.

> Conceptually, many authors thus have preferred to use verbal noun 'remembering' to underline the ongoing dynamic of construction and reconstruction and to replace the traditional idea of memory with one of being in a state of constant flux.[231]

This kind of language adjustment is an important part of the revisionary process—as we will see—but alone it is not sufficient for the rejection of problematic base assumptions. Where researchers still harbor computationalist assumptions, a change of language away from this computationalism may only bring confusion.[232] Brockmeier goes on to describe a "cultural paradigm shift" that has affected the way that human memory is regarded in literary, artistic, and media and technology fields, but explains that this shift has only been halfhearted among psychologists, and that it has been especially absent among those who work on the neurobiology of memory.

> But the general cultural paradigm shift has only to a very limited degree affected the conceptual shape of memory in this field, which is still very much defined in terms of storage, that is—to use the language of psychology and neuroscience—as encoding, retaining, and retrieval of information. This is all the more astonishing as there are many spectacular new findings and observations in this field that contradict and undermine the traditional view of memory.[233]

The language of storage and retrieval is still commonplace, and so are the concepts underlying these, despite the fact that so many empirical results sit so poorly within the framework

---

[231] Brockmeier 2010, p. 13.

[232] Morris Moscovitch writes that "memory does not exist until it is recovered" *because* "memory is a lasting, internal representation of a past event or experience that is reflected in thought or behavior" (Moscovitch 2007, p. 17). This tension between these claims is representative of many that are made by those scientists who are trying to move away from traditional models.

[233] Brockmeier 2010, p. 20. Another psychologist who is worried about the "continuing reliance" on the computer metaphor, William Randall, agrees with Brockmeier: "If one considers the language which discussions of memory routinely employ—the preponderance of references to 'encoding' and 'retrieval,' to the 'processing' of 'information'—one might think it were the only metaphor available; indeed, not metaphor, but fact" (Randall 2007, p. 611).

offered by these concepts.[234] This poor fit is resulting in some of the impasse found in memory science that was encountered in the last section. However, there have been developments that have at least attempted to grasp nearer the base when uprooting computationalist concepts. Two of these are worth investigating further. First, while the memory trace was once assumed to be localized in the brain, perhaps to a very specific neural configuration (or even a single cell), it is now accepted by the majority of researchers to be distributed across multiple elements or even regions. Second, there has been a shift away from the concept of storage and toward alternative concepts that better reflect what we have learned about the fluid nature of memory. In particular, a number of prominent memory scientists have elected to employ the concept of 'persistence' instead of the concept of 'storage' when characterizing the phase of the memory process between the encoding phase and the retrieval phase. Both of these are worth investigating because in each case the attempted conceptual revision has run into obstacles when the new concepts have been awkwardly juxtaposed with the old. It is also true of each of these revisions that they are grounded in many of the same considerations found above, and each would mark conceptual progress if only it could be freed from the trappings of the inherited concepts among which it is situated.

3.2.1 *Nonlocal traces and Long-Term Potentiation*

Writing in 1890, William James describes the question of the localization of cognitive functions in the brain as "the most stirring controversy in nerve-physiology which the present generation has seen."[235] One hundred and twenty-five years later, the only way in which this claim seems antiquated is the fact that it uses the term 'nerve-physiology'—otherwise the author

---

[234] Gallistel and King survey many points of this same poor fit, but come to the conclusion that memory scientists ought to relinquish these anti-computationalist sentiments and reject these revisions (Gallistel and King 2011).

[235] James 1890/2007, p. 30.

may have just as well been describing the *present* “present generation”. Localization of function is still a problem, and one to which we will return below. One localization issue that has seen some improvement in the last few decades is the localization of the memory trace. In part due to empirical results and in part due to research on artificial neural networks, most researchers have given up the notion that the memory trace is stored in one narrow and localizable spot. The “grandmother neuron”—the hypothesized cell that is solely responsible the memory and recognition of one’s grandmother—has largely, but not completely,[236] been consigned to the historical wastebasket.[237] Rather than being stored in one neuron, or even in one particular neural configuration, it can safely be said that the view that memories are stored in a distributed fashion is now the prevailing view.[238]

One of the reasons for this shift away from memory trace localization was the discovery of long-term potentiation, and the resultant emphasis on this process in research into the neurobiology of memory (and into cognitive neuroscience in general) in the 1980s and 1990s.[239] Long-term potentiation is simply the fact that postsynaptic nerve cells are stimulated more readily in the wake of repeated stimulation across the synapse in question.[240] That is, much like the “weights” in a connectionist network, the strength of the response of any given postsynaptic neuron to any given synaptic input will vary, and this will vary in part due to the frequency and recency with which

---

236 Waydo et al. 2006.

237 Gross 2002.

238 Sara 2000; Gross 2002.

239 Sara 2000, p. 78.

240 Strictly speaking, the dual processes of long-term potentiation and long-term depression—the latter being a reduction in responsiveness rather than a gain, resulting from different neurotransmitters—work together to provide useful changes in responsiveness. Without selective depression, aggregate LTP would tend toward unhelpful responsiveness thresholds.

that synapse has fired. Several memory scientists even equate the memory trace with this change in responsiveness: Howard Eichenbaum writes that "a change in field potential or cellular responsiveness that is persistent over many minutes or hours is identified as the putative trace of a memory mechanism."[241] Unless grandmother neurons have merely been replaced by grandmother synapses, of course, Eichenbaum must mean that a given *collection* of changes in field potential is identified as the memory trace. Thus the ability to later remember past events is grounded in the strengthening (or weakening) of certain connections in the neural network. This account was sketched *a priori* by William James a century before the biochemical mechanisms of LTP were under serious study:

> The currents, once in, must find a way out. In getting out they leave their traces in the paths which they take. The only thing they can do, in short, is to deepen old paths or to make new ones; and the whole plasticity of the brain sums itself up in two words when we call it an organ in which currents pouring in from the sense-organs make with extreme facility paths which do not easily disappear.[242]

Furthermore, James is here drawing on earlier themes in associationist proto-psychology of the nineteenth century, influenced by Hume a century earlier.[243] The discovery of long-term potentiation and its mechanisms are not the only recent developments to corroborate this view: we have also now designed and constructed artificial neural networks that operate in exactly this way. Connectionist networks can produce output which resembles past inputs. These networks accomplish this by the strengthening and weakening of responsiveness across collections of connections between nodes in the network. On the surface, these lessons have been duly absorbed

---

[241] Eichenbaum 2007, p. 194.

[242] James 1890/2007, p. 107.

[243] These ideas were influentially synthesized with brain science by Donald Hebb in the middle of the twentieth century, leading eventually to the study of LTP and LTD, as well as to our present understanding of artificial neural networks (Hebb 1949/2005).

by memory researchers: it is now widely accepted that memory storage is *distributed* in this fashion.[244] But revising the older notion of the engram, or memory trace, to suit these connectionist frameworks has produced an awkward juxtaposition: it is not clear that the mechanics of neural networks support the isomorphism and representation assumed to belong to the trace.

Connectionist representation remains a deeply embattled notion. Cognitive scientists still disagree as to whether or how neural networks can be said to be representational at all.[245] As James Garson notes, this disagreement has been "clouded by lack of clarity in defining what should count as the representational "vehicles" in distributed neural models."[246] These disputes are bound up with disagreements over the characterization of representation, and further clarification here will be important to several neighboring issues. However, we need not wait for the nature of connectionist representation to be determined to see how the storage of distributed memory traces in network models is problematic: the memory is usually treated as representational, but more importantly it is treated as structurally isomorphic to the past event it has encoded. Many of those who take connectionist networks to be representational, usually taking either an activation pattern or the sum collection of connection weights as a vehicle, admit that the internal structure of these vehicles has no features which can be mapped to that which the vehicle represents. William

---

244 Michaelian and Sutton (2013) trace the development of this acceptance. As they point out, this development has been tumultuous and multifaceted, and there is much present work being done still to stabilize and synthesize the resulting views.

245 This debate garnered significant attention during and following the "second-wave" connectionism of the late eighties and early nineties (Fodor and Pylyshyn 1988; Smolensky 1988; Elman 1991; Churchland 1993). Although the discussion has quieted since the mid-nineties, nobody believes that this is because the original controversies about connectionism and representation have been resolved; debates continue (Prinz 2006; Ramsey 2007, p. 145; Gallistel and King 2011).

246 Garson 2012, §6.

Ramsey imputes a certain sense of representation to connectionist network, but highlights the holistic quality of these representations:

> Thus, the smallest processing unit that lends itself to content ascription is thought to be the entire connection matrix, which is characterized as encoding the entirety of the system's stored information. For example, if the connections are thought to encode some set of propositions, then the most fine-grained semantic analysis that is possible is a holistic one that treats the entire connection configuration as representing the entire set of propositions.[247]

In at least some networks configurations, there are (slightly) more fine-grained semantic analyses available: Nicholas Shea presents a case for clusters, or regions of the networks activation state-space—the mathematical space representing all possible combinations of node activations—as the most appropriate candidate for representation.[248] These clusters, however, also lack internal features which are isomorphic to that which they represent. Whether there are interesting senses of 'representation' which allow for the total absence of structural isomorphism is a difficult question, but it is clearly the case that some types of representation do not require any isomorphism between the *internal* structure of the vehicle and the structure of that which is represented. Consider a penny used in place of a missing pawn in a chess game: no *intrinsic* features of the penny map to features of the pawn it represents. In this way Ramsey and others can claim that the distributed collection of connection weights is representational even though there are no finer-grained features of this representation that map to that which it represents. There are no second-order resemblances between the set of weights and the represented item. This conclusion does not enjoy consensus—there are still dissenters who seek meaningful, representational features in the features of the distributed representations—but so far no satisfactory account of this has been

---

[247] Ramsey 2007, p. 157.

[248] Shea 2007.

offered.[249] Whatever else connectionist representations may or may not be, they are not taken to demonstrate structural isomorphisms with those items they represent.

However, this leaves the mainstream account of memory with an inconsistent triad of commitments:

(1) Memory traces are encoded such that their features are isomorphic to the features of the past event.

(2) Memory traces are distributed (connectionist) representations.

(3) Distributed representations have no features isomorphic to that which they represent.

Memory scientists admit that traces are distributed but go on seeking the framework for the "transformation rules" that will allow us to understand the neural "code" in which they are written. As we have already seen, the failure to find such a code (or even plausible candidates) has pushed some researchers *downward* in their search, to look for the meaningful features of the trace at the molecular level.[250] But if those who study the properties of neural networks are not convinced any such features exist, memory scientists who invoke network models would do well to fully incorporate this absence of isomorphism. To talk of a "distributed trace" but also of a trace which is encoded and "read off" at retrieval is just to contravene our understanding of distributed representations in neural networks. Memory researchers have rejected the notion of a localized

---

[249] In one article Jesse Prinz attempts to identify meaningful features in the individual nodes, and thereby the dimensions in a vector in state-space (Prinz 2006b), but this seems to be mistaken in just the way Shea points out—conflating the semantic dimensions of a concept or representation with the vector dimensions of its vehicle (Shea 2007).

[250] For example, Lisman 2007 or Sweatt 2007 as cited above in 2.3.3 above. Not all move this direction, of course. Steve Peterson urges a broader and more inclusive neurobiological target when explaining remembering: "Thus, multiple forms of synaptic change in multiple locations may have to be combined to provide a satisfying explanation for associative strength measured in a particular set of behavioral conditions" (Peterson 2007, p. 51). Note that Peterson does not invoke the notion of a memory trace (and at an explanatory juncture where many would).

trace in name, but have retained assumptions about the nature of the trace and the remembering process that keep them from successfully actualizing this rejection. If memory scientists wish to retain (2) and (3)—which I will maintain that they have good reason to do, even though (3) is not often explicitly discussed in memory science[251]—then the assumption that memory requires an encoded, isomorphic trace must be revised or rejected.[252]

Charles Gallistel and Adam King, computationalists *par excellence*, have recently made very similar arguments, concluding that the isomorphic traces demanded by computational accounts of memory cannot be implanted by the distributed vehicles available in connectionist networks.[253] They explicitly disavow offering any alternatives, and they recognize that most cognitive neuroscientists endorse, at least vaguely, the view that traces are distributed elements of connectionist networks. Nonetheless, they are convinced that the isomorphism required by the storage and retrieval of memory is fundamentally at odds with the actual properties of both real and artificial neural networks. As a consequence of this (and their deeply engrained computationalism), they exhort cognitive scientists to reject even the possibility that memory traces are distributed, connectionist representations—which is just to appreciate the same inconsistent triad above, but to countenance the rejection of (2) instead of (1).

Even researchers who are sympathetic to warnings about the pitfalls of storage and who seek to revise those ill-conceived concepts we have inherited from classical computationalism

---

[251] Strictly speaking, I will ultimately suggest a revision to (2), since we will be doing away with "traces" altogether. A more general and mutually acceptable version of (2) might read: "The neural processes required for remembering deal in the currency of neural networks and connectionist vehicles."

[252] This is similar in spirit to Ramsey's rejection of "tacit" representation; Ramsey argues that the very notion of representation precludes the dispositional qualities of tacitness, and that any combination of these is confused (Ramsey 2007, p. 185).

[253] Gallistel and King 2011, especially chapters 11-14.

have a difficult time actually getting away from the stored memory trace that the archival conception of memory posits. In his book *Being There*, Andy Clark is a proponent of the lessons of connectionism and how these lessons can help us get out from under the spell of the computer metaphor. Clark explains one artificial neural network and how it does *not* have encoded, stored bits of computational "memory," but then he goes on to explain that its knowledge is stored in a different way.

> Most important, NETtalk's knowledge of text-to-phoneme transitions does not take the form of explicit symbol-string encodings of rules or principles. The knowledge is stored in a form suitable for direct use by a brain-like system: as weights or connections between idealized "neurons" or units.[254]

Clark would likely, if pressed, insist that 'stored knowledge' here means only the retained capacity to perform certain learned tasks, but this clarification is not trivial. The idea that information, knowledge, or memories can be "stored" in the weights of a network is part of what is keeping distributed memory research from succeeding in surpassing old models. In fact, this awkward use of the concept of storage is one of the many reasons for another class of recent revisions in memory science. Memory "storage" is quickly becoming obsolete in favor of concepts that fit the data better. Once again, this family of revisions stems from the appropriate attention paid to conceptual difficulties in memory science. However, like the attempt to distribute traces, the attempt to replace 'storage', *in situ*, with a more amenable concept, is hobbled by lingering allegiance to underlying computationalist assumptions.

---

[254] Clark 1998, p. 58. It is notable that in Clark's 2010 defense of his extended mind hypothesis, he continues to rely on memories as "stored information", but he admits that questions about the actual content of non-occurrent memory-beliefs are very difficult, and that these questions may "suggest the ultimate fragility of the very nature of intrinsic content" (Clark 2010, p. 11).

3.2.2 *Persistence and storage*[255]

In part due to many of the difficulties surveyed above, memory scientists have recently attempted to move away from archival models of memory in various ways. One target of criticism and development has been the traditional concept of *storage*. Yadin Dudai describes storage as a "misguided metaphor, of the type quite abundant in the science of memory."[256] Dudai is one of several cognitive scientists who have urged a conceptual shift toward *persistence* rather than storage.[257] That is, the phase of memory between the remembered event and the remembering event has been ill-represented, according to these critics, by the concept of storage. In light of all of the apparent difficulties caused by trying to reconcile the concept of 'storage' with the way that memory actually appears to work, it is easy to see why many researchers have sought alternative concepts. The notion of the "stored" trace has proven itself incompatible with various empirical findings, and models will be better served by characterizing the memory as *persisting* between the past event and the present recollection.[258] This dynamicism is one of the reasons Dudai calls storage a "misguided metaphor," and considerations like this clearly motivate a persistence approach. One can readily hypothesize that something persists without committing to a static, inert mode of being.

---

[255] An earlier version of this section was presented at the *European Conference on Cognitive Science* in Sofia in 2011, and subsequently appeared in collected proceedings (O'Loughlin 2011).

[256] Dudai 2007, p. 191.

[257] See also Asher Koriat's characterization of the "change in the dominant metaphor of memory" from storehouse to correspondence (Koriat 2007, 243).

[258] Howard Eichenbaum writes that ""persistence is an essential and defining feature of memory. Most simply stated, if a change in neural representation or behavioral expression does not persist, we do not call it memory" (Eichenbaum 2007, p. 193).

Another reason for the shift is that a persistent memory might enjoy various states of reconstruction, activation, or manifestation, whereas a stored memory seems to be either stored or retrieved. As it has turned out, many of the empirical results canvassed above support the former[259]—the activation states of memories seem to be anything but binary. There are almost no memory phenomena which sit easily within the framework of a stored, inert item that is activated when retrieved. These are all good reasons for memory scientists to have moved away from the concept of storage, but I will endeavor to show that the replacement of storage with persistence, while well-motivated, has been confused and half-hearted in practice.

One might wonder whether there can be any difference resulting from which of two terms memory scientists employ to describe a phase in their models. If our memory model is sufficiently rich, and accurate, can it matter what terms are used to denote its functional parts? Of course, the antecedent claim here is false: memory researchers are the first to admit that rich, accurate models are—at the moment—woefully absent. In the absence of complete models and understanding (and probably even in the presence of them), the language we use to characterize studied phenomena is a crucial component in our ability to overcome obstacles to our understanding of these phenomena. Dudai and his fellow memory scientists are right to be concerned that the language and concepts that have traditionally been brought to bear on the phenomena of memory may be impediments. Timothy Racine and Kathleen Slaney have recently edited a collection of essays on the use of conceptual analysis in psychology, citing in their introduction the fact that it "has proven surprisingly difficult for psychologists to find unanimous or even unambiguous answers to seemingly simple questions," and propounding Wittgensteinian conceptual analysis as an important part of the remedy.

---

[259] Loftus and Palmer 1974, Godden and Baddeley 1975, McDaniel and Masson 1985, Intraub et al. 1992, Roediger and McDermott 2000, Toth 2000.

> The root problem is the lack of consideration for the meanings of concepts that are in play in such work and the philosophical positions that are taken, explicitly or otherwise by the researchers who interpret such psychological terms in particular ways.[260]

In memory science, the choice of 'storage' or 'persistence' carries with it just these types of commitments and interpretations. In order to understand what is brought to a model by the concept of storage, and what is brought to a model by the concept of persistence, we must first compare these concepts in their natural habitat.

In consulting their respective uses in natural language, storage may at first appear to be a subset of persistence. That is, there may be many objects of persistence that may not be objects of storage—snowstorms or respiration, but also threats or inclinations. Surely, only in strained metaphor could any of these ongoing events, capacities or dispositions be said to be stored. This breadth exhibited by the array of phenomena that persist poses a difficulty for the cognitive scientist looking to apply this concept to memory. If it is the ability to remember that persists, then describing the middle phase of memory as persistence is only accurate in a tautologous fashion.[261] For any instance of remembering, it is an unassailable claim to assert that the ability to remember had persisted. With this generality comes an absence of theoretical commitments—but in some ways this is also helpful. The application of 'storage' to memory has borne too many implicit theoretical commitments, in part due to the specificity of the concept of storage.

Let us consider some common applications of storage: canned goods in a pantry, files in a filing cabinet, data on a hard drive. While in an artificial way any of these could be said to persist, none are the sorts of things of which persistence would be naturally predicated. Were you to ask if my stored canned goods persisted—while I would not outright reject the claim as I would the

---

[260] Racine and Slaney 2013, p. 1.

[261] Kathleen McDermott expresses similar concerns about broadening the concept of retrieval (McDermott 2007), and Daniel Schacter expresses a related worry concerning the concept of memory itself (Schacter 2007).

claim of a stored snowstorm—I would certainly pause quizzically. These cases of storage fall near the equivocal boundaries of the concept of persistence rather than squarely within them. Storage entails an inactive mode of being that goes beyond even the persistence of capacities or dispositions.[262] Storage also connotes a sense of localization. Whereas something might persist even while we cannot point to it—an attitude, a season, an economic trend—that something is stored entails that it is stored *somewhere*. The inherited connotations of language and metaphor can and do shape our theorizing in memory science.[263] Given this, it is clear that the choice of storage or persistence may present an explicit difference in hypothesis. A memory or memory trace that is stored is one that is inert, static, and probably in some form of neural code. When the stored item is activated, the consequence is a unitary manifestation. A memory or memory trace that persists is one that demonstrates "continual existence"[264] or "temporal extension"[265], but in a dynamic mode. The persistent item may manifest in varied ways, and in varied contexts.

The choice between storage and persistence yields a difference in the implicit commitments of the researchers' language, and perhaps a difference in the explicit commitments of researchers' hypotheses. In addition to these, there may also be a difference of *subject*. That is, we are predicating persistence or storage *of* something, but of what? Although this question sounds basic, the responses to it are fraught with controversy. Our folk psychological accounts have it that the memory itself is stored, or persists, but given that it is memory that we are seeking to analyze, this

---

[262]While capacities are in some sense inactive by nature, they are do not become inactive or active qua capacities. That is, a capacity which is activated is no longer a capacity. There is an ungainly redundancy, then, in applying the concept of storage to a capacity.

[263]See Roediger 1980 or Draaisma 2000 for a thorough and insightful exploration of just this thesis.

[264]Roediger et al. 2007, p. 191: "Persistence is continual existence."

[265]Eichenbaum 2007, p. 193: "Persistence is the temporal extension of a modification in neural representation or behavior resulting from experience."

is an unhelpful subject of storage and retrieval. Memory researchers, although they are united in rejecting "memories" as the subject of storage or persistence in the context of any precise explanation, offer a diverse selection of alternative subjects. Among others, neural firing patterns, changes in behavior, neural representations, learned information, memory representations, experience, and "brain changes" have each been described as the subject of persistence or storage.[266] If storage or persistence are given as technical terms that aim to designate concepts that describe a functional part of the memory process—and I take it that they are—then this disagreement as to the subject of the storage or persistence betrays a deep uncertainty in the application of these concepts.[267]

The problem of the subject will not fall as neatly into divisions, with the choice of storage or persistence, as did the previous considerations of linguistic connotations and differences of hypothesis. Many of the various candidates—trace, representation, neural pattern, experience—are susceptible to either characterization. Two outliers are worth mentioning. First, it has already been noted that capacities are appropriate subjects of persistence but not of storage. If it turns out that the capacity to manifest memory in some way is favored as a description of the item of memory, then persistence is favored. Second, an information processing view of cognition lends itself to storage rather than to persistence. If the subject of the middle phase of memory is best described as information, then storage will be the appropriate description of this phase. These respective terms bring inherited connotations of their own, along with what appears to be a ready-made difference in hypothesis. The persistent memory can be dynamic, mutable, and context-sensitive in all of the ways in which storage was found lacking. It could appear that memory

[266]See Eichenbaum 2007, p. 193; *Ibid*.; *Ibid*, p. 194; Roediger et al. 2007, p. 191; Mayes 2001, p. 191; Gabrieli 1998, p. 88; and Thompson 2007, p. 201, respectively.

[267]Tulving has similar concerns about the concept of memory itself (Tulving 2000, p. 36).

researchers have a new paradigm, albeit one that still needs some ironing out. In an examination of the concept of persistence as it applies to the neurobiology of remembering, John Lisman writes that "once memory molecules for persistence are identified, the investigation of persistence can be put on a solid footing."[268] Of course, it does not take much reflection on the search for "memory molecules of persistence" to start to wonder just what it is that is doing the persisting, and whether it is the sort of thing that *has* molecules. There are a number of confusions present in the way that the concept of persistence has been used, some of which can be illustrated by the wild divergences among researchers' characterizations of the subject of persistence, as surveyed above. This divergence belies the apparent unity in the newly forged persistence paradigm. Dudai's persistence of "learned information" is especially telling: persistence is still sandwiched between encoding and retrieval. Bearing these functional relationships, persistence necessarily ends up playing a very similar role to the one played by storage. In fact, as long as memory is characterized as something that is encoded during a past event, and then later retrieved, no word or concept for the phase between these will escape the problems of storage. Perhaps unknowingly accepting this parallelism, some neurobiologists have taken to using both persistence and storage, nearly interchangeably, perhaps as a way to hedge their theoretical bets.[269]

It is surprising that the concept of retrieval has retained its dominance even across this conceptual shift. We do not ordinarily *retrieve* something that is persisting. When Dudai, among others, employs persistence and retrieval together, he must mean at least one of these in a quite peculiar sense. The very fact that something has persisted usually means that it need not be retrieved, and the very fact that something needs to be retrieved implies that it has disappeared

[268] Lisman 2007, p. 206.

[269] See, e.g., Bekinschtein et al. 2006.

from some domain. There is a similar juxtaposition between persistence and encoding: only in a qualified sense can something that has been encoded be said to persist. Once something has been encoded, it lies dormant. If a thing is encoded and then later retrieved, it is difficult to see how it could be appropriately described as *persisting* in between these events.

My intent is not to hold science hostage to the foibles of natural language—if by ‘persistence’ these researchers mean to designate a technical concept that is amenable to encoding and retrieval, so be it. Rather, I aim to point out the equivocal nature of this adoption of the persistence models. A persistent item that is encoded and retrieved, for which a location and transformation rules are sought, “persists” in name only. If researchers find the evidence against storage moving, they ought to do more than change the label.. It is no surprise that researchers are having a difficult time establishing the “molecular basis” of memory persistence, or even establishing what would count as a subject of this persistence or its neural implementation, under the auspices of this confused half-rejection. ‘Retrieval’ implies a stored item in a location. ‘Encoding’ implies a trace, probably isomorphic to the past experience. Neither of these is suited to memory that *persists*, in the way that researchers who have abandoned ‘storage’ seek to capture with ‘persistence’. Understanding what a successful alternative model might look like will be difficult, but as a preliminary those underlying assumptions that are impeding this enterprise must be revealed. Only by stripping our concepts and models of these assumptions can we avoid implicitly committing researchers to unwieldy and contradictory amalgamations of concepts.

3.3 Computationalist assumptions, revisited

Each of the computationalist assumptions introduced in chapter two have caused difficulties—conceptual and empirical—for memory researchers. In the following section, I will revisit the computationalist assumptions outlined in section 2.3, illustrating how each has respectively contributed to the impasses, confusions, and inconsistencies above. Given that the assumptions of modularity, trace isomorphism, and hidden processes are not often made explicit or defended by those who study memory, the following considerations should at least provide motivation to investigate the need for and viability of these claims. This is especially so since so many within memory science or the philosophical study of memory admit that the conceptual frameworks brought to the subject stand in need of significant investigation. Additionally, revealing how these latent computationalist commitments are producing recalcitrant empirical difficulties will begin to undermine the assumptions themselves. Not only are these assumptions about modular, hidden processes and isomorphic traces deserving of scrutiny, there are also good reasons that each should be rejected.

3.3.1 *Modularity and independence*

While it is true that the modularity of memory systems, processes, and phases of memory processes is, in part, a product of computational assumptions about the mind and human remembering, there is also one motivation for modularity stemming from some of the very considerations above. The fact that memory is so fluid and multifaceted had induced researchers to posit different and modular memory systems long before other critiques of the traditional view came to light. Endel Tulving, in his influential 1983 book on episodic memory, immediately warns that he assumes that "there is no such thing as 'the memory'; rather, we have to distinguish between different forms of memory."[270] Tulving's work has been foundational for much of the work on

[270] Tulving 1983, p. 5. Nonetheless, many memory scientists (and philosophers) assume that there is an underlying unity, or overarching supertype, which subsumes these instances (e.g. Schacter 2007; see Michaelian 2011b for a philosophical treatment of this supposed unity).

memory systems, and it is easy to be sympathetic with his motivation. The realization that memory is not one thing is one that we will come back to with Wittgenstein in the following chapter. As we will see, Wittgenstein (as well as Ayer, Russell, Bergson, and for that matter Aristotle) agrees that there is a danger in too quickly assuming that memory is one particular process. The contemporary response to this in the cognitive science of memory, however, is to therefore assume that memory is *several* particular processes.

Tulving readily admits that these several particular memory processes often happen to hybridize in their natural environments. He understands the difficulty in separating the two as analogous to the difficulty encountered by any scientist who is devising clever experiments in order to isolate just one variable that is often, contingently, part of a natural compound.

> Given the close relation between the two systems, it is probably as difficult to find 'pure' episodic-memory tasks and 'pure' semantic tasks as it is to find sodium and chlorine as free elements in nature, although their compound, NaCl, is found in abundance. Most experienced and remembered events have factual contents whose characteristics are greatly influenced by semantic memory; performance on semantic tasks, on the other hand, may be influenced by episodic information.[271]

This much is surely right, and Tulving is also right to understand this difficulty as some evidence for epistemic modesty in claims about the taxonomy of memory systems. Following Tulving, we are hypothesizing about internal cognitive processes that are independent in principle but not in practice. Marcia K. Johnson makes a similar point, writing that "the subsystems are proposed to be modular in the sense that they can engage in some functions without reference to other subsystems, but multiple subsystems are normally operating and interacting in any complex task or situation."[272]

---

271 Tulving 1983, p. 55.
272 Johnson 2007, p. 60.

The assumption that remembering is a collection of modular processes grounded in distinct neural systems, which was above designated as horizontal componentiality, is the computationalist assumption which has been made most explicit in memory science. There have been, even since Tulving's work, occasional dissenters who have resisted the idea of distinct and relatively independent systems.[273] Because of this, the claim that remembering is made up of parallel alternative processes is sometimes made explicit and defended. Even in these cases, the inference to remembering as a parallel collection of internal, cognitive processes from the claim that memory is not a single internal cognitive process is not usually defended. This inference depends on the assumption that remembering is comprised of hidden, cognitive processes—an assumption to which we shall shortly return.

The assumption that remembering demonstrates vertical componentiality has not elicited this kind of attention: it is rarely made explicit or defended. Nonetheless, or perhaps as a cause of this, this componentiality is assumed by nearly every model of human remembering. Whether it be in terms of encoding, storage, and retrieval, or in terms of encoding, persistence, and retrieval (or others even more widely varied), each of these models assumes that there are component subprocesses that chain together to form the process or processes of remembering.

In the above considerations, we met with four classes of difficulties faced by memory researchers: two general classes of difficulties and two specific case studies of modern impasses in memory science. Each of these have been influenced by componentiality assumptions. The general difficulties were the following. First, the last fifty years of memory science have made it clear that, behaviorally, the way that memory actually operates is much more fluid and dynamic

[273] See, for example, most of the articles in Foster and Jelicic 1999 for a survey of dissent and revision among those studying the division of memory into "systems". This dissent can be traced back to more general claims from Lashley (1929).

than earlier models allowed. Remembering is deeply constructive, context-dependent, and multifaceted. Second, researchers studying memory at the behavioral level and researchers studying memory at the neurobiological level both seek and expect synthesis, but characterizing the relationship between the neural and the cognitive in any satisfying and rigorous way remains unsuccessful. In addition to these general difficulties, two specific impasses were introduced—the decades-long dispute over an appropriate characterization of amnesia and the (also decades-long, but not quite so many decades) dispute over how and whether to characterize nonhuman remembering processes in terms of the accepted taxonomies of memory.

It is easy to see how the assumptions of componentiality are making it difficult for modern models to adequately reflect the constructive and dynamic qualities of memory. One reason that the shift from storage to persistence has been stilted and unsuccessful is that persistence is taken to be a subprocess that neatly fits between the encoding subprocess and the retrieval subprocess. But as we have seen, the very concept of persistence does not sit well between these bookends. When it became clear that memory retrieval demonstrated varieties of instability that lessen over time, the intermediate phase of *consolidation* was introduced as a subprocess. Consolidation was, in the first instance, taken to follow encoding but to precede storage. Since remembering does not always involve an all-or-nothing retrieval after encoding, it was thought, there must be a period of instability, during which the process of consolidation is stabilizing the memories.[274] This introduction of a fourth phase or subprocess has generated considerable controversy, partly due to its different characterizations by different authors, partly due to its provenance: Joseph LeDoux writes that "consolidation is not something we subjectively experience or that has a distinct behavioral signature. As a result, the existence of consolidation is inferred rather than experienced

[274] See McGaugh 2000 or Sara 2007 for a survey of these developments.

or observed."[275] A consequence of this disagreement is that neurobiologists are no more certain of what would count as a neural process underlying consolidation than they are sure of what it would be to underlie encoding or retrieval.

Furthermore, behavior patterns in retrieval demonstrate curious instabilities throughout the life of a memory. Some of the data on positive and negative transfer, competences given impairments, priming, and retrieval-induced forgetting motivated researchers to posit multiple phases of consolidation or reconsolidation. Recently, there seems to be so much evidence of consolidation—intricate patterns of instability and stability across different contexts of retrieval—that some researchers are touting the ubiquity of these as vindication of the concept of consolidation[276] while others wonder whether the notion of consolidation is "sinking under its own weight."[277] The debates regarding how this additional subprocess can fit in alongside or between the subprocesses of encoding, storage, and retrieval. Of course, this worry is predicated on the assumption that the remembering process divides up into component subprocesses like encoding and retrieval. There is no doubt that the phenomena associated with consolidation and reconsolidation demonstrate an important *aspect* of remembering, but only on an assumption of componentiality do we thereby attempt to capture this aspect in a *subprocess*. Models that adequately account for memory's dynamicism have not been aided by the insistence that remembering is composed of inner subprocesses.

Componentiality assumptions have also plagued the search for neural correlates of memory. One localization puzzle facing neuroscientists of memory is what to make of brain

[275] LeDoux 2007, p. 171.

[276] For example, see Sara 2007.

[277] Dudai 2007, p. 180.

regions that appear to be correlated to memory, but which do not inhibit memory when lesioned.[278] Even the most basic and longstanding claim of neural underpinning, that the hippocampus acts as the "gateway" to memory by mediating retrieval has met with this result. Sara admitted at the turn of the 21st century that "more than 20 years later a direct test of this proposition was unable to provide supporting evidence,"[279] and exactly what can be said about the relationship between the hippocampus and particular memory functions is in no less dispute today.[280] But part of this difficulty stems from assumptions about modularity. Many brain systems appear to be functionally overloaded—some of the same areas of the brain that serve memory have been found to also serve imagination,[281] and some of the same brain regions associated with memory have been found to also serve language.[282] The idea that remembering is a collection of modular processes which are supported by or identical to a collection of modular neural processes drives cognitive neuroscientists studying memory to seek dedicated, modular neural processes that *are* remembering. The fact that such one-to-one correspondences continue to elude us applies pressure to the assumption that remembering is composed of modular processes.

Componentiality assumptions drive the framework of the amnesia puzzle. Horizontal componentiality plays some part in this, since memory competences performed by those with memory impairments comprise part of the evidence for the division of memory into memory systems in the first place. Some memory "processes" or "systems" seem to exist near the boundary

---

278 McDermott 2007, p. 229.

279 Sara 2000, p. 75.

280 Bartsch 2010, p. 2.

281 Schacter et al. 2012.

282 Ullman 2004.

between declarative and nondeclarative memory. One example of this is semantic priming. Patients with dense anterograde amnesias often retain procedural and implicit memory capacities.[283] This is the case even for semantic priming, changes in responsiveness to stimuli which share meaningful cognitive associations with earlier stimuli the amnesic patient does not report being able to recollect. Even normal rememberers exhibit this under the right conditions: we will all respond differently to something we have previously encountered, even if we do not recall encountering it, and even if the "it" in question is *semantic* rather than just sensory or syntactic.[284] One of the reasons amnesias are taken to never be complete amnesias is just this kind of amnesic competence: even the more dense amnesias are found to affect some "memory systems" and not others.

It is the assumption of vertical componentiality that frames the amnesia impasse, however. This can be seen above, in Squire's suggestion that the subprocess of storage and retrieval are simply conceptually inadequate for characterizing memory impairments. Specifically, he writes that amnesia would be best considered as a failure of "the mechanism that supports both storage and retrieval." The persistent incapacity to distinguish between encoding failure and retrieval failure just marks an incapacity to conceptually separate encoding and retrieval. For much of the amnesia debate, researchers have assumed that it made sense to ask whether encoding or retrieval had failed, and Squire is right to question the assumption of component subprocesses. In light of the admission that the assumption of separate component subprocesses is unwarranted in the case

---

283 The starting point for this research was established by Brenda Milner (Milner 1965) with the famous amnesic patient H.M., although this research was anticipated by the early twentieth century work of the Swiss neurologist Édouard Clarapède, who famously hid a pin in the palm of his hand while greeting an amnesic patient. Although she did not later recall the episode, she henceforth shied from shaking his hand.

284 A survey of this work is provided in Tulving and Schacter 1990. Semantic priming is usually accepted as a form of implicit memory, albeit of an exceptionally cognitive variety.

of amnesia, we might also attempt to establish which cases this assumption is warranted in. In fact, the consolidation and reconsolidation debates mentioned above show that even in normal remembering this assumption is problematic. The repeated or continuous reconsolidation researchers have recently been propounding highlights the essentially integrated character of “encoding” and “retrieval”.[285] Since encoding “processes” and retrieval “processes” are hypothesized cognitive processes that seem to elude both “behavioral signatures” and neural correlates, the fact that they provide impediments to understanding empirical results seems like good reasons to revisit them. If researchers were not tacitly committed to modular subprocess phases, there would be no puzzle about amnesia.

Lastly, the assumption of horizontal componentiality frames the puzzle about episodic memory and nonlinguistic creatures. The episodic memory system is taken to be modular and conceptually independent, which inclines researchers to limit hypotheses to the *presence or absence* of this system in scrub jays and other nonlinguistic creatures. The fact that the jays seem to have capacities for something that is a little like episodic memory and a little like procedural memory does not easily cohere with the assumption that memory systems are modular. Without the assumption that any given instance of remembering is *either* system x *or* system y, at least in principle, there would be no puzzle about the scrub jays or any other near-episodic cases among non-human animals. There are no puzzles about *what the jays can do*, only puzzles about which modular processes are being employed when they act. Tulving gets close to admitting this when

---

285 It briefly seemed like consolidation may help to resolve the amnesia impasse, but as Sara points out, introducing more component subprocesses only seems to have reiterated the same issues. “It is becoming increasingly evident as the literature grows that we are still stuck with the same old questions. Does the amnestic agent block consolidation, or now reconsolidation, or impair retrieval? Is the memory deficit permanent or is there spontaneous recovery or the possibility of recovering the memory by further treatments or reminders?” (Sara 2007, p. 186) Sara concludes from this tension between consolidation’s promise and the resistance of the amnesia problem that the amnesia impasse needs reframing.

he confesses that, in nature, the separate memory systems often hybridize. Like most researchers, however, Tulving continues to rely on the assumption that, in principle, these are separable processes. Tulving's hybrids are simply cases where both systems are engaged. Tulving no doubt seeks to remain faithful to the realization that there are in some cases very different ways of remembering, but the inference from there being different *ways* of remembering to there being separate underlying *processes* of remembering just relies on the assumption that memory is made up of underlying component processes.

None of these considerations are sufficient to completely undermine componentiality, but even any one of them alone is sufficient to motivate a more careful and explicit study of the claims in question. The assumption that memory systems capture independent processes that are alternative components that make up remembering is sometimes made explicit, but rarely defended. The assumption that the processes of memory are in turn made up of component subprocesses is almost never made explicit at all. Tulving and Squire get close to denying these assumptions, respectively, nonetheless—almost in spite of themselves. If any concepts in memory research deserve scrutiny, these componentiality claims must be good candidates.

3.3.2 *Isomorphism and memory traces*

The assumption that there is something in the brain that is structurally isomorphic to the past event or experience that is required for its future remembrance has undergone incredible strain under the present work in the study of memory. Memory scientists across different disciplines have realized that they do not know the nature of the memory trace. In many ways, the deep multidisciplinarity of the study of memory is responsible for this. This multidisciplinarity allows neurobiologists to carry on as usual, imagining that the trace whose neural correlate they seek has been defined by the psychologists while at the same time allowing psychologists to keep talking about the role that the trace plays and expecting that this trace will turn up in neuroscience any day now. Cognitive scientists talk about "translation rules" as though these are the key to what is still not known,[286] but of course no "translation rules" can be found before we have been able to specify what is being translated, and what it is being translated into. Only the assumption that there is this isomorphic neural configuration, or as O'Brien and Opie would have it, that there is a second-order resemblance between trace and past event, drives these mistakes. Let us examine what role this assumption plays in the impasses and difficulties cited above.

Chapter two demonstrated that the memory trace assumption has been informed and reinforced by computationalism. One of the traits imputed to traces that betrays a direct effect of this inheritance is the inert, static nature that is ascribed by default to the memory trace. The static nature of the storage phase does not enjoy consensus among researchers, especially with the advent of the controversial middle phase, or phases of consolidation and reconsolidation—these repeated or ongoing processes are sometimes said to have a hand in memory impairment[287]—but the trace isomorphism assumption is accepted by default in the characterizations of memory impairment. If a trace has been successfully laid down, then only a failure in the retrieval process will stop the result that the encoded trace is "read off" successfully. As we have already seen, these static or inert qualities of remembering have seen explicit rejection and attempted revision by many researchers in recent years. Researchers from many areas aim to construct models that are faithful to the dynamic, fluid properties canvassed above. The assumption that the memory trace's isomorphism with past events must be preserved in order for memory to operate is fundamentally at odds with this kind of dynamicism. The neural trace scientists are directed to seek becomes

[286] E.g. Dudai 2007, p. 79, as quoted above on p. 67.

[287] See Dudai 2006, although as Dudai admits reconsolidation is still controversial enough in almost every way that there are no widespread models of amnesia *based* on reconsolidation.

incredibly contorted under these pressures: this trace, or engram, must preserve isomorphisms across time that are also malleable and halting in many different contexts. Witness again David Sweatt's characterization of what he is looking for in the brain:

> Overall then, I conclude that molecular mechanisms involved in retrieval must be able to impinge directly upon the molecular basis of the engram, simultaneously triggering a re-activation of many of the same molecular changes involved in establishing the engram initially. Something very interesting is happening at the molecular level with retrieval—there seems to exist a specific mechanism capable of halting or erasing a chain reaction, but re-starting it in a plastic form that may end up being slightly different. *A mechanism that is capable of accomplishing this is quite mysterious to me based on present knowledge.* Nevertheless, having hopefully deduced its existence, the mechanism should be amenable to experimental investigation in the near future.[288]

Sweatt is right to find this mysterious. Indeed, Sweatt is simply explicitly thinking through the implications of what is all too often tacit in the science of memory. Given the assumption that the isomorphism of the trace is what allows memory to preserve past events, and the assumption that the trace is a neural record, then the dynamics found in the way memory actually operates must actually be reflected in the (mysterious) biochemical properties of the trace. This neural trace must, on these assumptions, have just the properties described by Sweatt. Of perception, Alva Noë writes, "This suggests that part of what has made the computational problem of vision such a difficult one is that it is framed in an artificially restrictive way."[289] The same could be said, *mutatis mutandis*, for memory.

An isomorphic trace is preservative by nature, and so any dynamics it apparently exhibits must be inconveniently incorporated into the trace model as a contingent feature. Much of the theoretical revision in memory science has followed this strategy, allowing the dynamicism of memory by hypothesizing consolidation phases or other flexibility generators which *attach* to the

---

[288] Sweatt 2007, p. 212, emphasis added.

[289] Noë 2004, p. 20.

basic encoding-storage-retrieval model. This is not the only way to revise these models: were it not for the assumption that the preservative, isomorphic qualities of the trace were essential to memory's operations, we might alternatively propose models for which the remembering process is essentially dynamic and constructive, rather than contingently so.[290] As long as our models are based on a framework inherited from machine memory, the isomorphic trace must remain inert at the core of the model.

Even Andy Clark, who once more offers good reason to be skeptical of the assumption that our minds operate in at all a similar manner to the objects of our metaphors, ends his warning not by denouncing the notion of an encoded trace, but by propounding revisions to what this "brain code" must be like:

> In retrospect, it is surely highly implausible that our brains (which are not so very different from those of some non-language-using creatures) should themselves use anything like the format favored by the thin projections of our thoughts onto public mediums like paper and air molecules. Brain codes must be active in a way in which text-style storage is not.[291]

Clark is certainly right to be wary of any model in which the similarities between brain and metaphor are too convenient, but in framing the operation of memory in terms of "stored" "brain codes", he is condemning the resultant models to the very commitments he seeks to undermine. This is an example of the way that the isomorphism of the trace is implicit in characterizations of memory, even among those researchers who are neither computationalist nor lacking in philosophical acumen.

The assumption of an isomorphic trace probably only plays a minor role in the puzzle about nonlinguistic episodic memory—trying to understand the scrub jays ability sometimes means

[290] Kourken Michaelian's "Generative memory" emphasizes the non-contingent nature of construction and consolidation in human memory, and attempts to frame a model accordingly (Michaelian 2011a). We will return to Michaelian's framework below.

[291] Clark 1998, p. 58.

asking whether the jays have recorded the "what, when, and where" of the caching episode[292]—but this assumption certainly plays a major role in the amnesia puzzle. It is just the assumption that all and only successful remembering must issue from a successfully preserved trace that assures researchers that memory impairment must stem from *either* a problem with the encoding of the trace or a problem with the retrieval of the trace. If construction and dynamicism are essential to remembering, and there is no supposition that memory's preservation of the past can only be possible through the preservation of an isomorphic trace, then the amnesia puzzle, as it is presently formulated, will dissolve.

3.3.3 *Remembering processes and forgetting processes*

A long recognized difficulty for the study of memory is that although memory is taken to be a compound process involving earlier and later phases before the actual remembering event, memory is studied only through retrieval or the absence of retrieval. Susan Sara spells this out in her article on the neurobiology of remembering: "Memory lends itself to study through its retrieval whether it is evaluated by the behavior of a mouse in a swimming pool, a verbal report from a human subject, or inferred from an electrophysiological event."[293] Given that the precursor processes that are posited, at the cognitive level, are those that are supposed to map onto *neural* processes that underlie them, there is a special difficulty in establishing any physiological processes at all as the physiological processes that underlie memory. Those neural processes that

---

[292] Clayton and Dickinson 1998, p. 274. This characterization of episodic (or episodic-like) memory is curiously ambiguous, and probably only depends on the trace assumption on one of its two readings. It is not always made clear whether the rememberer must be able to *recall* each of these consciously (but we surely enjoy some episodic memories we cannot place in time or space), or whether these must constitute part of the recorded information in the trace, whether recalled or no (in which case these do not seem to provide a sufficient condition, since procedural or semantic memories must also issue from traces that contain this information).

[293] Sara 2000, p. 73.

are supposed to serve remembering are always a process away from the phenomena we study. Sara gives this illuminating description of the problematic corner researchers have backed themselves into by virtue of this assumption of a double process:

> Retrieval must somehow involve initial activation of relevant intrinsic networks, selection of relevant extrinsic stimuli, and integration of these different sources of information into a meaningful trace. … The role of subtle, but significant, environmental stimuli in triggering these processes is intuitively obvious, and has been investigated systematically in animals and humans. Nevertheless, virtually nothing is known about the physiological processes underlying the act of remembering. The initial process must involve some orientation of attention to a particular stimulus or ensemble of stimuli. How those particular stimuli are recognized as "meaningful" or how they can activate the specific distributed network presumed to be the neuronal substrate of the memory still remains unknown.[294]

Like the contortions undergone by the neural target of Sweatt's investigation, the hidden cognitive process of remembering has also taken on mysterious shapes from the pressures informing it on standard models. Retrieval, in the double-process model, is a cognitive process which yields certain behavioral symptoms (under certain conditions) and is implemented by a particular neural process. The cognitive process must, as Sara confesses, perform miraculous feats on this view: it must recognize (but not "remember", at risk of regress) meaningful stimuli and match these stimuli to relevant traces (but again, without remembering or already having the trace handy to compare).[295] Once more, it is no wonder that researchers are puzzled by the specific functionalities of their investigative target. Revising memory models to account for the dynamicism found in remembering behavior cannot accommodate the miraculous function needed by this inner process.

The conjecture that there is an inner, cognitive process of storage and retrieval is also one of the impediments to researchers being able to understand and adequately characterize amnesia.

---

[294] *Ibid*., p. 76.

[295] These difficulties will be returned to, at length, in chapter four.

Famously, the amnesic patient H.M. was thought—like all amnesic patients—to have encoding difficulties, until Warrington and Weiskrantz showed that his behavior was modified by experiences that he failed to consciously recall.[296] Researchers have struggled to describe this non-conscious memory in terms of its underlying processes. This was not a struggle to identify the physiological processes that were and were not functioning adequately for H.M.—this we knew, with comparative precision. This was also not merely a struggle to understand what H.M.'s learning abilities actually were and under what conditions—discovering this was taken to be step toward understanding the inner processes. Lastly, this was also not a struggle to understand the inner processes of H.M.'s remembering as in the way that H.M. experienced remembering. Indeed, questions regarding the *phenomenology* of amnesia rarely play much of a role in investigating the hidden cognitive processes. Rather, the attempt to characterize H.M.'s processes, as well those of all memory impairments, hang up on the attempt to understand where in the *hidden* process the failure has manifest.

A related phenomenon is known as infantile or childhood amnesia—the fact that we do not remember our early childhood. It is sometimes referred to as the *paradox* of infantile amnesia, because our experiences in our early childhood clearly have important and lasting psychological effects. At least one of the proposed resolutions to this paradox is widespread retrieval failure.[297] Of course, the same back-and-forth which plagues the amnesia puzzle generally also undermines the debate about a satisfactory account of childhood amnesia. Since encoding and retrieval are essentially hidden, it is not clear what distinguishes retrieval failure from encoding failure, even in principle. This is just what Hardt and his colleagues were trying to resolve by testing impairment

[296]Warrington and Weiskrantz 1968.

[297] Spear 2007.

in mice, and the resilient failure to distinguish hidden subprocesses is just what Matzel and Miller criticize in the Hardt et al. experiments. Matzel and Miller reveal how the Hardt et al. experiment must assume the very correlation—between subprocess and behavior—that the experiment is supposed to test.[298]

The problem that Matzel and Miller point out in the Hardt et al. experiment and interpretation is indicative of deep-running difficulties in the attempt to explain memory impairment in terms of the underlying component processes of storage and retrieval. The question of whether memory impairment is a matter of storage failure or retrieval failure cannot be answered in the way that Hardt et al. hope to answer it. Matzel and Miller conclude their commentary with an exhortation to revisit the *question*, however, rather than the answer, taking the failure of the Hardt et al. strategy not as a call to try a different approach to deciding between storage and retrieval, but instead as undermining the nature of the storage/retrieval debate.[299] After reflecting on the decades of debate, Larry Squire also backs off from assumptions about the mental subprocesses that purportedly underlie memory and memory impairment. In the passage above, Squire urged us to think of amnesia as a loss of whatever it was that supports "both storage and retrieval." Although neither party quite advocates an explicit reassessment of the more general claim that memory is a hidden cognitive process, the claim that the hidden processes researchers have been dealing in no longer track the phenomena provides good reason to re-examine our notions of hidden processes altogether. Indeed, these rejections of component mental processes as

---

298 "These complications are precisely analogous to the one asserted by Hardt et al. to have plagued previous efforts to distinguish storage from retrieval interpretations of experimentally induced amnesia" (Matzel and Miller 2009, p. 671).

299"More important is the very nature of this debate. A question that might better be asked is under what circumstances do each of these interacting influences prevail, and what might be the functional consequences for memory processing?" (Matzel and Miller 2009, p. 671).

viable explanations of the phenomena of amnesia contain echoes of some of the Wittgensteinian critiques of mental subprocesses we will meet in the next chapter.

Lastly, the assumption that memory is a hidden process or collection of hidden processes drives the very shape of the debate concerning nonlinguistic rememberers. There is no puzzle regarding what the jays can and cannot do overtly—this is easily observed. The puzzle is born from the supposition that episodic remembering is a hidden process that may or may not exist "in the mind" of the rememberer. Once more, this isn't even taken to be a phenomenological process (at least not usually—this distinction is not always made clear). That is, the puzzle is not simply about what the jays are experiencing—were that the case, this puzzle would hardly belong to memory science proper. What memory scientists cannot agree on is whether there is an inner, cognitive process of episodic remembering which has been demonstrated by the jays' caching behavior. Again it seems as though any evidence for the existence of this process can be reinterpreted as evidence for the lack of episodic memory on the jays' part, and researchers are explicitly concerned that it may be impossible *in principle* to decide whether nonlinguistic creatures have episodic systems. If this has not already aroused investigators suspicions as to the nature of these hidden processes, it should.

None of the above considerations decisively undermine the assumptions of componentiality, isomorphism, or hidden processes. Many of these considerations from recent memory science, however, provide good reason to be suspicious of these assumptions. These empirical considerations are intended to expose and destabilize these deeply-held commitments, to show that the careful philosophical investigation of these assumptions is long overdue, even if the assumptions themselves can at first glance appear benign. Some memory scientists have already begun this effort—the recent compilation of essay written *about* concepts in memory research *by* memory scientists is a salient example of this[300]—but fortunately a number of philosophers have already examined these assumptions. As other scholars have already noted[301], some of the criticisms of modern cognitive science coming from modern cognitive scientists are asymptotically approaching some of the criticisms offered by Wittgenstein and other philosophers early in the twentieth century.

[300] Roediger et al. 2007. This work followed from a much earlier article by Tulving exhorting memory scientists to this task (Tulving 1991), and it has influenced even more recent work such as Brockmeier 2010.

[301] Moyal-Sharrock 2009; Boncompagni 2013; Sutton 2014.

CHAPTER IV

PHILOSOPHICAL DIFFICULTIES FOR ARCHIVAL MODELS

> Many of the conceptual confusions fundamental to modern cognitivist theory had already been identified and widely recognized before the 'cognitive revolution' of the 1960s. Yet, whenever such confusions are pointed out, they are either fleetingly acknowledged, only to be quickly forgotten, or, more usually, emphatically denied.
>
> Alan Costall[302]

> Why should I deny that there is a mental process? But "There has just taken place in me the mental process of remembering…." Means nothing more than: "I have just remembered….". To deny the mental process would mean to deny the remembering; to deny that anyone ever remembers anything.
>
> Ludwig Wittgenstein[303]

4.1 Wittgenstein on memory

As we have already seen to some extent, many of Wittgenstein's remarks on mental processes, understanding, psychology, or language directly bear on the study of memory. Even if these were all we had, Wittgenstein's approach to mind and cognition would still be an important tool for resolving conceptual confusions in the philosophy and sciences of memory. As it happens, however, there are also many remarks from Wittgenstein that directly consider the nature of memory. Although it would be too strong to claim that these remarks constitute a systematic or complete Wittgensteinian *account* of memory, there are recurring Wittgensteinian themes concerning memory that provide powerful critiques of some of the common ways we tend to automatically conceive of remembering. Very generally, Wittgenstein often directs our attention to all of the ways in which memory *does not* seem to be an inner process of storing and retrieving.

---

302 *The Unconscious Theory in Modern Cognitivism*, p. 312.

303 *Philosophical Investigations* §306.

Danièle Moyal-Sharrock credits Wittgenstein with having a significant destabilizing effect on these assumptions in the cognitive science of memory, and characterizes the Wittgensteinian contribution thus:

> Wittgenstein's major contribution to the elucidation of the concept of memory is his discrediting the picture of memory as information storage and boldly replacing it with the idea that memory is nothing but an ability and that, in some contexts, remembering amounts to a way of acting; that is, to an act or expression which does not result from introspection or retrospection.[304]

Moyal-Sharrock summarizes several themes present in Wittgenstein's writings on memory. In keeping with his approach to many subjects, Wittgenstein draws our attention to what actually happens in particular cases of remembering, and demonstrates several things by means of these examples. Instances of remembering take many forms, as Wittgenstein repeatedly indicates. Some of the properties often ascribed to memory (and characterized as necessary for remembering) are simply not to be found in some of these forms. In our characterizations of many phenomena, especially mental phenomena, Wittgenstein cautions against neglecting particular cases in favor of presuppositions of unity and analysis. This is especially tempting when our language happens to treat certain psychological terms as though they each refer to an entity or process. Michel ter Hark usefully calls this view *referentialism*, according to which "the primary role of psychological words is to stand for or refer to things, properties and processes." Ter Hark is right that "referentialism is one of the main targets of Wittgenstein's philosophy of psychology."[305] Many of Wittgenstein's remarks on memory fit just this description.

Indeed, the surface grammar of the language of memory and remembering is beguiling. 'I have just remembered' has the same apparent structure as 'I have just eaten' or 'I have just had a

[304] Moyal-Sharrock 2009, p. 218.

[305] Ter Hark 2000, p. 196.

headache'. 'My memory of that is faint' has the same apparent structure as 'My stomach is full' or 'My pain is subsiding'. These apparent similarities can be deceptive. The referentialist move is to take 'remembered' as though it must refer, like 'eaten' or 'had a headache', to a particular kind of process that has just happened. Similarly the referentialist takes 'memory' to refer to a particular kind of entity. Taking these terms to refer to processes and entities of some kind is just the "first step" that Wittgenstein calls "decisive moment in the conjuring trick."[306] Once we have begun to treat memories and remembering as entities and processes that have details that can be further investigated, we begin to investigate as though the details will be revealed. These entities and processes do not seem to be physical (such as stomachs or eating), and neither do they seem to be straightforwardly phenomenological (such as pains). Given that we are already treating them as processes whose details can be revealed by investigation, these processes and entities can only be *hidden* processes, and these "memories" can only be hidden memory traces.

Of course, to diagnose how we may have come unjustifiably to a belief is not to demonstrate that the belief itself is false.[307] The way that the language of remembering can mislead us runs parallel to the way that over-reliance on the computer metaphor can mislead us: each can explain why certain assumptions remain tacit, or why certain assumptions resist revision even under adverse conditions, but neither provides sufficient grounds for rejecting these assumptions. Many of Wittgenstein's remarks on mental processes, and on memory in particular, sketch diagnoses of how we have gone astray, but many others demonstrate problems, not just with

[306] Philosophical Investigations §308. The full passage is quoted above, in 3.1.1, and it is notable that it ends with "And now it looks as if we had denied mental processes. And naturally we don't want to deny them."

[307] Although Wittgenstein does remark on the importance of retracing our steps in order to free ourselves of the philosophical puzzles these steps have brought us into, famously writing that philosophy "unravels the knots in our thinking; hence its results must be simple, but its activity is as complicated as the knots that it unravels" (*Philosophical Occasions*, p. 183).

the route, but with the destination. Several of Wittgenstein's remarks on remembering and mental processes are directly concerned with revealing difficulties in the mainstream view that remembering is an inner process that retrieves something that is stored. The subjects of these criticisms can be usefully divided into three targets. First and most generally, Wittgenstein seeks to demonstrate the fundamental diversity among the phenomena of memory. One way to show that our usual models are inadequate is simply to consider those cases that these models do not easily cover, and to show that there is a broad variety of cases like this. Second and most specifically, Wittgenstein questions the need for a stored memory trace. Last, Wittgenstein asks us to the very notion of hidden processes, and what a hidden mental process of remembering could be. Cognitive-level processes are crucial for mainstream models of remembering, and there are Wittgensteinian considerations—some of which echo and presage considerations by other philosophers—that tell against both the need and the plausibility of inner, cognitive-level processes of remembering.

4.1.1 *Ways of remembering*

Wittgenstein insists that that remembering can take many forms. He draws our attention to the many ways we can be rightly said to remember, and to the sometimes surprising differences among these ways of remembering. He is, of course, not the first to declare that remembering can proceed in distinct ways—in the introduction to the history of memory systems above, we encountered a collection of thinkers, before and after Wittgenstein, who remarked on the multiplicity of the forms of remembering. This collection includes nearly every researcher who is working on memory in the present day, if the thesis in question is only that remembering can take different forms. However, Wittgenstein's characterization of the plurality we find in the phenomena of remembering—and more generally, in the phenomena of thought—is subtle, and has been influential. Norman Malcolm remarks on the problematic inclination on the part of philosophers to ascribe fundamental unity to their subjects, and on the way that this inclination has played out in the study of memory.

> It is generally true that when a philosopher enters into the study of a concept (such as knowledge, belief, intention, memory) he assumes that the concept has a *unity* which can be disclosed by "analysis." In the case of the concept of memory there are a multitude of linguistic forms and locutions in which the noun "memory" and the verb "remember" are employed. The philosopher's normal assumption will incline him to suppose that one of these linguistic forms is preeminent, basic, or central, and that the other forms are offshoots from the central one.[308]

Malcolm is right that philosophers have sometimes tended, perhaps too quickly, to try to subsume problematic particular cases under a single category that best describes the "core" of remembering.[309] We will return to some examples of this later in this chapter. He is also right to draw this lesson, broadly speaking, from many of Wittgenstein's remarks. However, this lesson about the disunity of memory has very often been duly absorbed by memory researchers in many disciplines. In fact, as we have seen above, the notion that memory is comprised of multiple, distinct systems enjoys remarkable consensus.[310] The question of whether these do or do not belong under one naturally distinct supercategory is taken to be an open one.[311] If the only Wittgensteinian lesson regarding differences among particular cases of remembering were one of disunity, there may not be much more here for memory studies to garner from Wittgenstein's work.

---

[308] Malcolm 1977, p. 16, original emphasis.

[309] Bergson, who helpfully distinguished "habit-memory", or procedural memory, from other kinds of remembering—and whose distinction was an early and important one—referred to other kinds of memory as "true remembering" (Bergson 1896/2004).

[310] See Squire 2004 for a review and analysis of mainstream characterizations of memory systems in the sciences of memory.

[311] See Michaelian 2010 for a philosophical treatment of this question. Michaelian ultimately answers in the negative, but does not assume anything about the unitary status ahead of argument.

Wittgenstein's remarks on the many ways in which we remember are valuable as a starting point for a certain critique of the way memory is often characterized, but this critique emphasizes both difference and similarity among diverse instances of remembering. This work has been followed up by other philosophers and researchers, but the multiplicity of memory that we find focused on in Wittgenstein's work exhibits subtleties underappreciated by much contemporary research, even though, as we have seen above, contemporary memory researchers readily admit that remembering consists in a great variety of processes and manifestations.

A naïve approach to memory may induce us to think that to remember is to bring an image of the past to mind. Several philosophers, many of whom were directly or indirectly influenced by Wittgenstein, have explicitly rejected this picture of memory. In the follow extended passage from *The Concept of Mind*, Ryle even names such a view "the stock account" of remembering:

> The stock accounts given of reminiscence give the impression that when a person recalls an episode belonging to his own past history, the details of the episode must come back to him in imagery. He must 'see' the details 'in his mind's eye', or 'hear' them 'in his head'. But there is no 'must' about it. If a concert-goer wishes to recollect just how the violinist misplayed a certain piece, he may whistle the bungled tune, or play it on his own fiddle just as the artist had done it; and, if her repeats the mistake faithfully, he is certainly recollecting the artist's error. This might be his only way of recalling how the artist had gone wrong, since he may be poor at going over tunes in his head. Similarly a good mimic might recapture the preacher's gestures and grimaces only by reproducing them with this own hands and on his own face, since he may be poor at seeing things in his mind's eye. Or a good draughtsman may fail to recollect the lines and the rigging of a yacht, until he is given a pencil with which to delineate them on paper. If their mimicries and delineations are good and if, when they go wrong, their authors duly correct them without being prompted, their companions will be satisfied that they have recollected what they had seen, without desiring any additional information about the vividness, copiousness or connectedness of their visual imagery or even about its existence.[312]

Ryle is surely right that visual imagery is no necessary phenomenological part of a process of remembering. He is also right to emphasize the ways in which our remembering can be embodied and nonvisual; as predominantly visual and intellectual creatures, we tend to automatically attend

---

312 Ryle 1949, p. 250. Cf. Wittgenstein, *Philosophical Investigations* §450.

to those examples, not just of memory but of many cognitive capacities, which neatly conform to a visual-intellectual framework. Wittgenstein also reminds us that "the most varied things may happen" in remembering, and cautions us against assuming that an image is necessary ahead of actually studying the phenomena. [313] Moreover, he directly argues against the necessity for an image.[314] However, the assumption that visual imagery is necessary for remembering is one that has been, by and large, successfully purged by memory researchers. Contemporary memory scientists certainly recognize that not all instances of remembering need involve experienced images. Of the three major memory systems—procedural, semantic, and episodic—two do not even *usually* involve images. Procedural memory is expressed in the successful performance of activities that require learned skill, and usually has nothing to do with imagery. Semantic memory—the ability to remember learned facts—may involve imagery in some cases, but often involves the recollection of non-visual facts that are not suited to any relevant images. Even episodic memory is not always taken to require experienced images, even though it is unanimously admitted that it *may* involve such images. Hence the first point from Ryle, that not all memory need require experienced images, is one that that is already recognized in memory science.[315]

The passage from Ryle's *Concept of Mind* is not only dispelling a need for visual imagery, though. Ryle has also here claimed that some remembering processes *consist in* embodied behaviors, absent any experienced "inner" phenomena at all, visual or otherwise. It is not clear

313 Philosophical Grammar §42; E.g., Zettel §25: "If I say "I saw a chair in this room", I can mostly recall the particular visual impression only very roughly, nor does it have any importance in most cases." Cf. Zettel §650.

314 See, for example, *Blue and Brown Books* p. 89.

315 This is not to say that Ryle's point was or is unnecessary in general: it is probably the case that many non-specialists outside of memory research still casually depict remembering as the calling up of an inner image. Bernecker characterizes calls this view of memory "representative realism", and he is probably right to claim that it is still influential in philosophy of mind and epistemology (Bernecker 2009, p. 32-33).

that Ryle himself separates these points in this passage: he characterizes the view he rejects as holding that the rememberer "must 'see' the details 'in his mind's eye', or 'hear' them 'in his head'," and yet he emphasizes that what is not necessary is the *visual* image. These are two separate theses: that visual imagery is not necessary for remembering, and that there are no experienced phenomena necessary for remembering. The latter is more contentious and interesting. That there is no essential phenomenology of remembering is not always recognized even by memory researchers.

A similar point was made a few years later by A. J. Ayer. In *The Problem of Knowledge*, Ayer considers and rejects the experienced image as a necessary component of remembering.[316] He then considers Russell's suggestion from *The Analysis of Mind* that there is a phenomenological marker of familiarity that essentially accompanies remembering. Ultimately, Ayer also rejects this as a requirement.

> Perhaps the correct answer is that there is no one thing that is universally present in every such instance of remembering. Sometimes it is a matter of one's having an especially vivid image; sometimes, with or without an image, there is a feeling of familiarity; sometimes there is no specific mental occurrence: it is simply then a matter of one's seriously saying 'Yes, I do remember'. There can, indeed, be said to be distinctive memory-experiences, in the sense that remembering an event, whatever form it takes, 'feels different; from merely imagining it, or believing that it occurred. But these experiences do not essentially consist in the presence of a special sort of object.[317]

Although in his previous chapter (on perception), Ayer had used considerable force explicitly disagreeing with Ryle on several points, his exposition of the phenomenology of memory makes Ayer sound distinctly Rylean. Perhaps, as we will see, it would be more accurate (or at least more

---

[316] Ayer 1956, p. 138. It is notable that Ayer devotes one of his five chapters in *The Problem of Knowledge* to memory, and that he worries that memory is too quickly assumed to be a unitary process parallel to perception.

[317] Ayer 1956, pp. 146-147.

fruitful[318]) to describe both Ryle and Ayer as sounding distinctly Wittgensteinian. Ayer is clearer than Ryle that what he does not find essential to remembering is not just experienced visual or imagery phenomena, but *any* particular experienced phenomena at all. There is no phenomenological process or state that is part of what it is to remember.

Even this may seem to have already been absorbed by the memory science community. Different memory systems are taken to demonstrate different experienced phenomena. After all, the densely anterograde amnesics who successfully demonstrate skill learning despite being unable to recall ever previously performing the skill in question—like the case of H.M., as told above—do not experience any "memory phenomena" while engaged in procedural remembering. Still, Ayer's examples (and Ryle's, and Wittgenstein's) are not limited to procedural or skill memory. These instances of remembering are declarative, and some are clearly examples of what would today be called episodic memory. The phenomenological point here is that even *episodic* remembering requires no particular experienced state or process. Again this may not seem worth emphasizing in the science of memory—the encoding, storage, and retrieval processes that are taken to be essential components of remembering are not usually characterized as *experienced* processes. It seems uninterestingly true that we can remember without experiencing anything in particular. Still, this is an important lesson not just for philosophers or non-specialists who might slip into treating memory as though it must be a particular experienced process, but also for memory researchers who are positing the states and processes of memory. It is not always made clear what level these putative processes can be found on, and it can only help conceptual

---

318 Detailing or defending the threads and extent of historical influence among these is not my aim. Ryle is certainly sometimes described as categorically "belonging to the Wittgensteinian tradition" (ter Hark 2000, p. 195), Ayer is usually not described so. For the purposes of this chapter, I will focus on Wittgensteinian considerations about the received view of memory, which Ryle and Ayer (whether by direct or indirect influence, or simply independently) are also, in part, expressing.

confusions in memory science to make this clear. The "inner" processes of remembering are taken to be infra-behavioral but supra-neural. This is a difficult level to work with. Experienced, phenomenal states and processes are also sometimes taken to be infra-behavioral and supra-neural, and it is easy to mistake any talk of a level between behavior and brains as talk of phenomenology. However, standard models of memory do not offer these processes as experienced processes. If the putative component processes of memory are neither behavioral nor neural, *nor phenomenal*, so be it, but keeping this characterization in full view will be important to resolving the difficulties that the models of these processes have yielded.

These remarks do shed light on the nature of the "inner" processes of remembering, but this nature will be the subject of the next section. Although Ayer is clearer than Ryle in distinguishing between the hypothesis that remembering requires inner, visual imagery and the more general hypothesis that remembering requires a particular sort of inner, experienced state or marker, there are also two distinct hypotheses intertwined in Ayer's comments. Ayer concludes that there is no "special sort of object" whose presence is required in any process that can rightly be called remembering. But his earlier claim that "sometimes there is no specific mental occurrence: it is simply then a matter of one's seriously saying 'Yes, I do remember'" seems to be asserting something much stronger. This might be put in terms of a simple scope ambiguity: to deny that all remembering requires some mental occurrence may just be to deny that there is any one occurrence that all remembering possesses (Ayer's explicit conclusion in this paragraph), or it may be to deny whether each instance of remembering need contain some mental occurrence at all—whether there are cases of remembering that do not involve any mental phenomena at all (a view attested to in this paragraph and elsewhere in the chapter by Ayer[319]). That is, there may be cases of remembering that are constituted by actions like "seriously saying 'yes, I do remember'."

Wittgenstein's remarks on memory also dispense with imagery, and with any particular mental process as a necessary component. Adding to these, however, Wittgenstein also takes up this last and strongest rejection: the rejection of any inner processes at all being necessary for remembering. There are ways of remembering that leave nothing hidden. There are versions of this argument against the need for inner processes of remembering in several of Wittgenstein's later writings, but the following passage from *Philosophical Grammar* is worth quoting at length:

> What happens when I remember the meaning of a word? I see before me an object of a certain colour and I say "this book is brown and I have always called this colour 'brown'". What sort of act of remembering must take place for me to be able to say that? This question could be put in a much more general form. For instance, if someone asked me "have you ever before seen the table at which you are now sitting?" I would answer "yes, I have seen it countless times". And if I were pressed I would say "I have sat at it every day for months". — What act or acts of remembering occur in such a case? After all I don't see myself in my mind's eye "sitting at this table very day for months". And yet I say that I remember that I've done so, and I can later corroborate it in various ways. Last summer too, for example, I was living in this room. But how do I know that? Do I see it in my mind's eye? No. In this case what does the remembering consist of? If I as it were hunt for the basis of the memory, isolated pictures of my earlier sojourn surface in my mind; but even so they don't have, say, a date written into them. And even before they've surfaced and before I've called any particular evidence into my mind, I can say truly that I remember that I lived here for months and saw this table. Remembering, then, isn't at all the mental process that one imagines at first sight. If I say, rightly, "I remember it" the most *varied* things may happen; perhaps even just that I say it. And when I here say "rightly" of course I'm not laying down what the right and wrong use of the expression is; on the contrary I'm just describing the actual use.[320]

Wittgenstein asks "What sort of act of remembering must take place for me to be able to say that?" and then proceeds to show that this question is misguided, that there is no act of remembering that must take place, and that there are some cases in which remembering is

[319] See especially Ayer 1956, p. 137 and p. 141-142. The latter will also be discussed in further detail below.

[320] *Philosophical Grammar* §42, original emphasis.

exhausted by acting or speaking.[321] The most varied things may constitute remembering. There are cases—and many of these are garden variety, unexceptional instances of remembering—where what there is to see in the *mental* process of remembering is already in full view. This is not yet to say anything about neural processes, of course. Nor is it to claim that nothing is ever hidden from view—neither Ryle nor Wittgenstein countenance this kind of naïve behaviorism. Rather it is simply to establish the existence of some cases of remembering where what mental process was to be found was already the one before us.

> Think of putting your hand up in school. Need you have rehearsed the answer silently to yourself, in order to have the right to put your hand up? And *what* must have gone on inside you?—Nothing need have. But it is important that you usually know an answer when you put your hand up; and that is the criterion for one's *understanding* of putting one's hand up.
> Nothing need have gone on in you; and yet you would be remarkable if on such occasions you never had anything to report about what went on in you.[322]

That we later can give explanations of our raised hand is no guarantee that these explanations track prior deliberations or inner processes.[323] Remembering in this case *just is* raising your hand. If there are cases like this, or like the earlier example of assenting to have seen this table countless times, or to have lived in this house last summer, then it is not the case that remembering requires any inner mental processes at all. Ayer agrees. In considering a very similar situation, he writes of someone who has been asked about the past that "their remembering just consists in their getting

---

321 To think that the utterance "I remember sitting here" must be primarily a report on an inner process (that could in principle be true or false) rather than the expression of memory itself is just what Austin warns against, cautioning that philosophers are prone to understanding a sentence's most basic job as a descriptive one (Austin 1975, p. 72).

322 *Zettel* §136, original emphasis.

323 A similar point concerning intention was influentially made by Anscombe: "One might not have a 'mind' to do something, distinguishable from uttering the words" (1957, p. 4).

the answer right. Whether they are helped to do so by conjuring up images, or consciously delving into their past experience, is irrelevant."[324]

Once more there may be an ambiguity here. Wittgenstein seems to be making the strong statement that nothing need have gone on inside the rememberer, and in so claiming he seems to also preclude *unconscious* cognitive processes, not just experienced mental processes. Indeed, Wittgenstein warns us about the dangers of positing unexplored unconscious processes in a number of contexts. This also implies, of course, the weaker thesis that some acts of remembering include no *experienced* processes over and above the experienced process of, say, speaking. The stronger thesis, however, is that there are some acts of remembering that include no processes at all, as constituent parts, other than the processes of say, speaking or raising one's hand. This is not to say that there are not neurobiological processes that are necessary for remembering: but these necessary physical processes are no constituent part of what it is to remember.

Wittgenstein asks us to consider actual cases of remembering, and notes that these are sometimes as simple as raising your hand or stating that you remember. If this is so, he has demonstrated that a search for *the* process of remembering is misguided.[325] The contemporary memory scientist may still at this point feel like she is in accord with this lesson, since memory has been carved up into systems: it is taken as an open question whether there are any memory processes common to all of these systems.[326] Furthermore, the 21st century memory scientist

---

324 Ayer 1956, p. 137.

325 Citing Wittgenstein, Ramsey makes a similar claim concerning representation, which he suggests is a "cluster concept", rendering the search in cognitive science for "the" definition of representation moot (Ramsey 2007, 8). Remembering seems to be at least as good a candidate as representation (if not better) for this family-resemblance treatment.

326 This is discussed by memory scientists in Squire 2004 and Johnson 2007 to some extent, but investigating underlying commonalities among the systems more often falls to philosophers of memory, e.g. Michaelian 2011.

might just lump cases like the raising of one's hand or even the rote answering of a question into procedural memory; after all, verbal responses are sometimes trained in a manner quite parallel to other skill training. This is an interesting response, and one which deserves (and below shall receive) more attention, but it does not fully account for Wittgenstein's considerations about the variety in the forms remembering can take.

Wittgenstein often brings our attention to spectra or collections of interrelated cases of mental phenomena. One such example is from *The Blue Book*.[327] Although Wittgenstein is only indirectly addressing remembering in this case, it illustrates the way that he uses a collection of mental phenomena to show, not just the differences among the cases, but also very particular similarities.

> Let us consider the process of estimating a length by the eye: It is extremely important that you should realise that there are a great many different processes which we call "estimating by eye".
> Consider these cases:--
> Someone asks, "How did you estimate the height of this building?" I answer: "It has four storeys; I suppose each storey is about fifteen feet high; so it must be about sixty feet."
> In another case: "I roughly know what a yard at that distance looks like; so it must be about four yards long."
> Or again: "I can imagine a tall man reaching to about this point; so it must be about six feet above the ground."
> Or: "I don't know; it just looks like a yard."
> This last case is likely to puzzle us. If you ask "what happened in this case when the man estimated the length?" the correct answer may be: "he *looked* at the thing and *said* 'it looks one yard long'." This may be all that happened.[328]

Wittgenstein does not construct these collections of cases merely to demonstrate that there is one among them that has no need of hidden processes or explanations. Rather, these collections also draw our attention to similarities among these respective mental phenomena. In the fourth case, we seem to already be at the bottom of the explanation—the estimation of length was just what

---

327 Another example, from the *Brown Book* (BBB p. 85), will be discussed below.

328 *Blue and Brown Books*, p. 11, original emphasis.

was before us: the man *looked* at the thing and *spoke*. If we look again at the second or third case, it is apparent that a similar bottom level is not far away. If we ask the man in the second case how he knows what a yard at that distance looks like, he may well give an answer much like the man from the fourth case. If we ask the man in the third case how he knows how tall to imagine a tall man, his answer may also be similar. Even in the first case, certain questions will eventually elicit the answer, "I just see it". Furthermore, we could go on constructing an indefinite number of cases in between any two of these. Each has similarities to the other, and each has differences from the other as well. Cases like the fourth are not particularly unique, they just lay bare aspects of cognition that are often masked.

In an example just following the above in *The Blue Book*, Wittgenstein asks us to consider someone who is told to choose a yellow ball from among several. We may be inclined to think that the subject must "bring a yellow patch to mind" in order to find the yellow ball, but then Wittgenstein asks us to consider the case where the subject was told to "imagine a yellow patch". Surely, Wittgenstein remarks, he need not imagine a yellow patch first, in order to successfully follow the order. Again, cases where nothing is hidden are not special, they only make plain what can otherwise be obscured. Remembering can consist in many things, a great variety of processes. Some of these will leave nothing hidden, others will happen to include, for example, inner, visualized phenomena. These differences are important, but it is also important that any two of these cases belong to a chain of interrelated intermediate instances.

But if this is the case, then although there is not one process category to which all remembering belongs, neither are there *several* process categories to which cases of remembering belong. It is true that some rote memorization of verbal responses looks a lot like procedural memory, like the learning of skills. However, this does not demonstrate that some apparently verbal memories actually reside in the procedural "system"; rather, it demonstrates that the boundary between these apparent categories cannot divide systems. I may raise my hand in response to a question, and that may be all that happens. I may bring an image to mind, concentrate on this image and study it, and then raise my hand as a result. Or I may commit any number of actions that comprise the intermediate cases between these. Whether there happen to be inner images involved is a comparatively unimportant feature of these processes. Ayer comes to the same conclusion. He admits that a memory image can prompt one to "form an accurate belief about one's past experience," but insists that we can also "form such beliefs without the assistance of an image." He concludes that there seems to be no very sharp difference between procedural (what he and other philosophers tended to call 'habit memory') and episodic memory (what he calls 'event memory').[329]

If these philosophers are right, then the intermingling of memory systems is more than the product of contingent biological facts, as suggested above by Tulving. It is not just that the memory systems usually collaborate, or are frequently coincidental. Rather, whether an instance of remembering is an instance of episodic or procedural memory may in some cases be an essentially indeterminate matter. We can certainly distinguish subgroups of remembering cases by identifying features or criteria, but we should not expect these to resolve into distinct systems. Michel ter Hark writes that the core of Wittgenstein's approach in his writings on the philosophy of psychology is just this: "to accept and to describe psychological indeterminacy as it is,"[330] rather than to give in to our inclinations to categorize psychological processes sharply. We do

329 Ayer 1956, p. 142. Although in part stemming from Ryle, it is noteworthy that this lack of a sharp distinction between these ways of remembering may have problematic implications for the distinction between *knowing*-how and *knowing*-that—a distinction that, rightly or wrongly, is often attributed to Ryle (E.g. Stanley and Williamson 2001, p. 411).

330 Ter Hark 2000, p. 193.

"need to remember that the process of thinking may be *very various*"[331], but we also need to remember that that the complexities among these varieties of thinking are no less rich and intricate than the complexities of the lives to which they belong. "We find that what connects all the cases…is a vast number of overlapping similarities, and as soon as we see this, we feel no longer compelled to say that there must be some one feature common to them all."[332] This density of this variety precludes, rather than implies, sharp psychological categorization.[333]

4.1.2 *Must there be some stored thing?*

Wittgenstein's work will ultimately hold reasons to be suspicious of an inner, cognitive states, processes, and entities as constituents of remembering. Before getting to these more general concerns, there is also one very specific hypothesis in memory to which Wittgenstein directed his attention: the hypothesis that remembering requires an encoded, stored trace in the brain. This, of course, just is the computationalist assumption isolated above, the assumption that the apparent isomorphism between the remembered event and the remembering phenomenon requires an isomorphic trace to mediate between them. Wittgenstein's criticisms of this assumption are among those of his remarks that most straightforwardly address subjects that belong to 21st century cognitive science. Consider again the passage from *Zettel* that preceded chapter one:

> I saw this man years ago: now I have seen him again, I recognize him, I remember his name. And why does there have to be a cause of this remembering in my nervous system? Why must something or other, whatever it may be, be stored up there in any form? Why must a trace have been left behind? Why should there not be a psychological regularity to which

---

331 *Zettel* §62, original emphasis.

332 *Blue and Brown Books*, p. 87. The cases characterized by Wittgenstein in this passage from the Brown Book are cases of *comparing*, utilizing memory in particular.

333 Moyal-Sharrock draws some similar lessons from Wittgenstein regarding the ways that we remember, but then goes on to advocate for the exclusion of implicit or nondeclarative remembering from what we rightly call "memory" (Moyal-Sharrock 2009). This division and exclusion seems to be in direct contrast to the "overlapping cases" lessons from Wittgenstein.

> no physiological regularity corresponds? If this upsets our concept of causality then it is high time it was upset (*Zettel*, §610).

This might seem out of step with Wittgenstein's usual subjects and methods. Throughout much of his work, Wittgenstein characterizes philosophy in a way that emphasizes differences from the scientific enterprise. He warns against the tendency to make philosophy more scientific, and seems to consider the proper subjects of these respective endeavors quite distinct.[334] Why, it might be asked, does he in these remarks suddenly take up what is or is not going on in the brain as his subject?

The claim that there is a certain configuration or structure in the brain is an empirical one, and it is a proper subject of scientific inquiry. However, when this claim is assumed to be true ahead of empirical inquiry, when it is assumed that is must be true given the nature of mind and causation, the boundary dividing scientific inquiries from philosophical inquiry has been crossed. The trace, as such, is not something that has been found in the brain, or might have been found in the brain, which scientists afterward named and characterizes. Rather, the memory trace is taken as a necessity, even though scientists are not quite sure how to look for it. In taking it thus, it is the scientists who are unwittingly doing philosophy, not the philosophers who are unwittingly doing science.

Furthermore, Wittgenstein has good reason to characterize the processes of remembering (or thinking, or expecting, or reading) as they pertain to what goes on in the brain. The basic psychological indeterminacy, as ter Hark calls it, of these mental processes is in tension with the

---

[334] *Philosophical Investigations* §109, which emphasizes the nonscientific character of philosophy, is one well-known remark indicative of this attitude. The Wittgensteinian emphasis on the separation of philosophy and science is lauded by some who seek to apply Wittgensteinian insights to cognitive science (Boncompagni 2013, p. 30; Hacker 2013) and it is criticized by others who seek the same (Sutton 2014). The extent to which Wittgenstein himself may have de-emphasized this separation, especially in his last writings, remains controversial (Marconi 2012).

notion that any of them are realized or implemented by a specific, encoded brain process.[335] That is, one of the reasons we resist accepting this indeterminacy as it is, according to this Wittgensteinian view, is just that we already believe we understand these processes as having a particular and hidden nature. We cannot help but to imagine that the indeterminacy is epistemic, and that the processes are determinate in the mind, or at least in the brain. The more certain we are that remembering is tracked by a particular brain process, the more certain we shall be that remembering cannot be all that various.

In denying the neural trace, or even in denying an underlying neural process that implements remembering, Wittgenstein is not denying that all manners of neural activity may be necessary conditions for any particular case of remembering. That is, for any process or event that is a case of remembering, there are many particular neural conditions necessary for that event to have taken place. Additionally, there may be (and very probably are) some general neural process- and state-types which are necessary conditions for most or all instances of remembering. We could even allow for neural process-types which are physically necessary and sufficient for remembering (though the Wittgensteinian view of memory will be hard to reconcile with this in other ways) while still denying an isomorphic trace. Danièle Moyal-Sharrock writes:

> Wittgenstein is not denying that we need brain activity for thought, and in that limited sense, remembering is causally dependent on the brain– on brain structures in different brain areas and on synaptic modifications in these areas – but it does not follow that these structures are representations of particular memories, stored and encoded in the brain.[336]

To assume that there is a neural trace of a past event stored up in the nervous system, which is activated upon actual remembering of the event, is to assume more than a merely causal

---

[335] Ter Hark 2000.

[336] Moyal-Sharrock 2009, p. 215.

connection between brains and behaviors. That the brain changed as a result of the past event is one claim. That certain brain changes are necessary for particular remembering behaviors and phenomena is another. That these changes include a physiological regularity that corresponds with the psychological regularity is yet a third—this is the claim that there is a stored trace which maps to the past experience. As Moyal-Sharrock notes, Wittgenstein is not very concerned with prognosticating about what we will and won't actually find in the brain—rather, he wants to point out that traces that map to particular memories do not follow from the mere causal links between the physiological and the psychological. Even when Wittgenstein expresses the most confidence in the non-trace picture of memory and brain-mind correlations, he expresses this in terms of what is *possible*:

> Indeed, I confess, nothing seems more possible to me than that people some day will come to the definite opinion that there is no copy in either the physiological or the nervous systems which corresponds to a *particular* thought, or a *particular* idea, or memory.[337]

It is this possibility that Wittgenstein's remarks are supposed to draw out, which is the same possibility precluded by the assumption in modern memory science that a corresponding, isomorphic trace can be individuated in the brain. On the rare occasions that memory scientists, or philosophers, explicitly defend the very notion of a trace, they tend to do so on the grounds of the apparent causation, or "retained information", across time and between past event and remembering phenomenon.[338] This is just the leap that Wittgenstein asks us to consider more closely. Take, for example, the following extended passage from *Zettel*.

> Imagine the following phenomenon. If I want someone to take note of a text that I recite to him, so that he can repeat it to me later, I have to give him paper and pencil; while I am speaking he makes lines, marks, on the paper; if he has to reproduce the text later he follows those marks with his eyes and recites the text. But I assume that what he has jotted down is not *writing*, it is not connected by rules with the words of the text; yet without these jottings

---

[337] *Last writings on the Philosophy of Psychology*, §504, original emphasis.

[338] Sutton 1998, p. 303; Tulving 2007, p. 66; Bernecker 2009, p. 5.

> he is unable to reproduce the text; and if anything in it is altered, if part of it is destroyed, he sticks in his 'reading' or recites the text uncertainly or carelessly, or cannot find the words at all.--This *can* be imagined!--What I called jottings would not be a *rendering* of the text, not so to speak a translation with another symbolism. The text would not be *stored up* in his jottings. And why should it be stored up in our nervous system?[339]

In this example, Wittgenstein has clearly admitted that the jottings have a causal relationship with what is remembered. Not only are they necessary conditions for proper reading of the text, but there are relatively tight connections between the jottings and the reading, such that varying the former in subtle ways can also vary the latter in subtle ways. Nonetheless, the jottings do not match the text. There are no rules of translation that will convert the jottings to the text; rather, the only available conversion of jottings to text is the particular individual in this particular context. By stipulation, the jottings are "not connected by rules with the words of the text." It is worth emphasizing that this can be imagined, to demonstrate that tight causal relationships do not guarantee isomorphism. The text as it was recited to the subject bears an isomorphic relationship with the later recitation by the subject. We do sometimes preserve such isomorphisms by encoding the features of the first event in some physical transcription, but Wittgenstein claims that this is not the only way these can be preserved. There is no reason to assume that this must be how such preservation operates in our nervous system, and there is nothing metaphysically exceptional is the supposition that these psychological regularities may not correspond with any identifiable physiological regularities, even in principle:

> The prejudice in favour of psycho-physical parallelism is also a fruit of the primitive conception of grammar. For when one admits a causality between psychological phenomena, which is not mediated physiologically, one fancies that in doing so one is making an admission of the existence of a soul alongside the body, a ghostly mental nature.[340]

---

339 *Zettel* §612, original emphasis.

340 *Remarks on the Philosophy of Psychology*, vol. 1, §906.

Indeed, this is revealed by those who defend the trace by simply citing causation or physicalism. If one is convinced that a psychological regularity at t_1 that does not map to a physical regularity at time t_0 must instead map to a non-physical regularity at t_0, then one will be equally convinced that physicalism entails a memory trace.[341] After all, humans can remember the past even when the past is separated from the present by periods of unconsciousness; so if anything that can be individuated at t_1 must map to some cause that can be individuated at every prior time between regularities, then these mapped causes during unconscious periods must be physiological. But this is confused: causation does not logically imply continuous regularity. The jottings might enjoy a causal relationship to both the heard and recited versions of the text, and yet there may be no regularities identifiable, even in principle, in the jottings. To declare this is not to declare a non-physical cause, or any kind of "ghostly mental nature", it is simply to point out logical possibilities among corresponding features that are causally related across time.

It may at this point still seem like such possibilities are not physically possible, that the way causation must work in this world really does guarantee continuous regularity and identifiable isomorphic causes. An alternative account, which allows for causation across a physical medium without any isomorphic structural regularities in the interim, will be elaborated on below. The Wittgensteinian point in question at present is that if we can imagine psychological regularities without physiological regularities, then the memory trace is no forgone conclusion. The concepts of correspondence and causation do not require this kind of physiological mediation.

4.1.3 *The myth of inner processes*

[341] I take it that this is the reason Michaelian considers it safe to assume the existence of a trace until proven otherwise (Michaelian 2011a, p. 331), following Sutton (1998, p. 303) and Bernecker (2009, p. 5), since such an assumption would be otherwise quite curious.

The last collection of Wittgensteinian insights that directly bear on the study of memory is also the collection which is most widespread throughout his work, coming into contact with several common Wittgensteinian themes about mind. These are his remarks on the very notion of "inner processes", and on "inner processes" of memory in particular. Throughout the Wittgenstein corpus there are recurring arguments against the need for, and the viability of, inner mental processes in a variety of explanations and contexts. "In philosophy one is in constant danger of producing a myth of symbolism," Wittgenstein writes in Zettel, "or a myth of mental processes. Instead of simply saying what anyone knows and must admit." The *Philosophical Investigations* opens with the story of a grocer who is choosing five red apples by comparing items and properties in the world with the purported images in his mind. The private language argument is an attack on something like an inner process of ostension. The collected remarks that were until recently called part II of the *Philosophical Investigations* often revolve around the putatively inner processes of aspect-seeing. Throughout his work, Wittgenstein warned about the "myth of mental processes" and the inclination to classify psychological phenomena as such.

He was also explicitly concerned with the processes of memory. In *The Brown Book*, Wittgenstein tells iterations of a story involving remembering a color for the sake of comparing it to another color on a bolt of material.[342] In the first iteration, the remembering subject holds an image of the color in his mind, compares this image with various colors in the real world, and chooses the appropriate real-world color. In the second iteration, he has no memory image before his mind, but rather only feels "a sort of mental tension" when looking at the various colors on offer, which relaxes when he reaches the desired color. In the third iteration, the subject "goes to the shelf without a memory image, looks at five bolts one after the other, takes the fifth bolt from

342 This version is in the *Blue and Brown Books* (p. 85), but there are similar stories told in the *Investigations*, as well as the *Remarks on the Philosophy of Psychology*.

the shelf." In this last case, we might be tempted to respond with Wittgenstein's interlocutor: '*But this can't be all comparing consists in*,' —Wittgenstein diagnoses our worries thus:

> When we call these three preceding cases, cases of comparing from memory, we feel that their description is in a sense unsatisfactory, or incomplete. We are inclined to say that the description has left out the essential feature of such a process and given us accessory features only.[343]

However, argues Wittgenstein, if a person can be trained to produce an image in response to a cue or an order, automatically as it were, why can we not be trained to produce an action in response to a cue, to simply choose the right fabric? If remembering can just be raising one's hand, or choosing the right fabric, then the addition of an inner process is superfluous. We are inclined to think that the essential features of the remembering process have yet to be revealed, but any process we imagine must bottom out at automatic responses somewhere. Given that this is so, an inner remembering process that gets started automatically, proceeds through some subconscious inner subprocesses, and then initiates experienced mental phenomena and action has been rendered superfluous.

Many have recognized that this argument against the success of inner, mental mechanisms is germane to current theorizing in psychology and cognitive science.[344] In a recent article, Daniel Hutto emphasizes the need for what he calls "elucidatory philosophy" when dealing with our characterizations of psychological phenomena. Although he is here focused on an entirely different family of underlying mental mechanisms supposed to explain surface phenomena, Hutto could just as easily be discussing explanations of amnesia that invoke component mental processes:

---

343 *Ibid*, p. 86.

344 For example, versions of this argument and its applications in cognitive science appear in Bennett and Hacker 2003, Noë 2004, Hutto and Myin 2013, and Boncompagni 2013.

> If we think about folk psychology...it is clear that there need be no underlying theory that serves as its basis. Rather, oft-posited "deeply tacit" rules and principles that we can identify are really just those embedded in—immanent within—the activity itself. Although we can formulate and codify them—upon reflection and to some limited extent—this in no way shows that these rules exist independently (in any sense). If so, it is not the rules that structure and explain the activity itself, rather they are revealed by it. However natural, it is a mistake to suppose that they must be, somehow, intellectually grasped and then applied if we are to account for the structure and content of the relevant activity. … Indeed, there is an unhealthy, indeed circular, relationship between such theories and the evidence used to support them.[345]

Hutto's article is concerned with applying these insights to "theory of mind" characterizations of folk psychological explanation, but the Wittgensteinian insights he urges us to apply in this domain are easily applied to the mental subprocesses that are said to underlie remembering. Wittgenstein addresses the assumption that "special processes" provide the underlying basis for remembering, pointing out that such processes are certainly not part of the ordinary phenomena of memory:

> One might almost marvel that one can answer the question "What did you do this morning?" – without looking up historical traces of activity or the like. Yes; I answer, and I wouldn't even know that this was only possible through a special mental process, remembering, if I were not told so.[346]

We do not posit special mental processes because these are a part of our experience of remembering. Rather, these are products of our habitual impulse to theorize. Wittgenstein asks us, repeatedly, to focus on what actually happens in mental phenomena, and to do away with any theorizing assumptions about the processes that underlie what happens. He mentions the notion of a stored memory trace, in particular, several times, asking us "Why must something or other, whatever it may be, be stored up there *in any form*? Why *must* a trace have been left behind?"[347] "Must I assume that if someone draws or describes or imitates something from memory, he reads

---

[345] Hutto 2009, p. 209.

[346] *Remarks on the Philosophy of Psychology*, *vol. 1*, §106.

[347] *Zettel* §610, original emphasis.

off his representations from something or other?!—What supports this?"[348] For Wittgenstein, the urge to posit a process of storage and retrieval beneath the phenomena of remembering as we know them is just what we ought to resist.

These underlying processes are tempting, in part, because the search for underlying mechanisms that help to explain experienced phenomena sometimes proves successful. Acid reflux underlies heartburn; decreased circulation underlies numbness in a limb. The problem is that storage and retrieval have to play both and neither of these roles: these special processes are neither experienced nor straightforwardly biological. Cognitive neuroscientists presently seek the biological processes that underlie storage and retrieval, and psychologists seek to explain the phenomena of remembering and failing to remember in terms of these processes. This intermediate position engenders a certain slippage: in neurobiological contexts storage and retrieval can be treated like heartburn—a high-level process for which we need only identify the biological underpinnings—and in psychological contexts storage and retrieval can be treated like acid reflux—the underlying mechanism that explains the everyday phenomena. This equivocal characterization makes resisting these special processes difficult. Only a sufficiently comprehensive perspective will expose the conjuring move as such: storage and retrieval cannot properly belong to either of these categories, each must permanently be an "uncomprehended process in an unexplored medium."[349]

If this is the case, it should come as no surprise that difficulties have cropped up in isolating and analyzing these subprocesses and their behavioral and neural correlates. We have decided that remembering is a matter of storing and retrieving, and that we will later figure out the nature this

[348] *Zettel* §34.

[349] *Philosophical Investigations* §308.

storage and retrieval. We then have trouble identifying the neurobiological mechanisms that constitute or underlie these mid-level processes, and we have trouble identifying what kind of mechanisms these are supposed to be. We are then unable to decide which process—storage or retrieval—has failed in any given instance of behavioral impairment. But this is, of course, just what we would expect if these “deeply tacit” rules—in Hutto's words—did not underlie the phenomena at all—if the immanent patterns in remembering and forgetting were the only mental processes to be had.

We ought to resist these subprocesses in part because they are not a part of our experience; they are unquestioned theoretical assumptions, and as such their pedigree is suspect. We also ought to resist because these mental subprocesses would not help to explain the phenomena of remembering even if the remembering process needed such explaining. This is demonstrated in the amnesia impasse above. Carrying a stored mental item around with us won't allow us to match this item up with the world we experience anyway,[350] and a retrieval subprocess, as Hardt et al. make clear, would need an on-board storage system in order to effectively underwrite remembering. It is far from clear how an explanation of amnesia in terms of storage and retrieval could even be laid out, such that these components of the process were conceptually distinct. Although Wittgenstein’s remarks draw our attention to the phenomena of successful remembering rather than failed remembering, the impulse to explain memory impairment in terms of these underlying subprocesses stems from entrenched assumptions about the component processes that ground everyday memory. As such, the difficulties that arise for these explanations of impairment can illuminate just what is problematic in these entrenched assumptions.

[350]Cf. *Philosophical Investigations* §604, §605.

Some of this rejection of inner processes can be seen in the attitudes and revisions of 21st century memory scientists. Above, we were met with Larry Squire urging memory scientists to dispose of the operation or failure of subprocesses of storage and retrieval when explaining memory impairment, and instead to characterize such impairment in terms of the whole process. While Squire rejects this particular set of mental processes (storage and retrieval) as unhelpful in an explanation of this particular remembering phenomenon (impairment), the generality of either of these can be varied to yield several distinct ways to reject inner, component processes.

Indeed, in many of Wittgenstein's remarks it is clear that he questions the efficacy of storage, in particular, as helpfully explanatory for *any* of the phenomena of remembering. This is stronger than Squire's conclusion: an inner "storage" component is not just useless for explaining memory impairment, it is useless for explaining memory at all. While Squire does not generalize to this stance, we have already seen many examples of other memory scientists questioning the role of storage in just this way. As it turns out, there are several empirically-based reasons for doubting whether the storage and retrieval models track any of the phenomena of memory very accurately at all. We might reject storage and retrieval as satisfactory component processes in an explanation of memory impairment or we might reject storage and retrieval as satisfactory component processes in an explanation of memory more generally.

More generally, we might reject the inclination to explain amnesia in terms of inner processes altogether. Matzel and Miller might be espousing this when they urge us not to ask about which component process fails in memory impairment. Part of the value in applying Hutto's "elucidatory philosophy" to these assumptions is the ability to understand the degrees of freedom permitted by the interplay of the assumptions in question. We can reject storage and retrieval as satisfactory component processes in an explanation of memory impairment, or we can reject inner, component processes (or the failure of these) altogether in an explanation of memory impairment. In this case, storage and retrieval, *per se*, are not the problem—the problem is in trying to *go beneath* the failure of remembering at all. Under certain conditions, the subject remembers in certain ways; under other conditions, the subject fails to remember in certain ways. Perhaps we can study the causes of, the conditions for, and manifestations of memory impairment without studying the processes beneath memory impairment.

Alternatively, we might generalize both of these. Wittgenstein seems to reject the need for component mental processes altogether in explanations of memory phenomena. In this case, the very idea of inner, mental processes posited to explain the familiar phenomena of remembering is misguided. Our reliance on such processes is a product of over-theorization, and will ultimately impede our efforts to understand memory. Many researchers in psychology and cognitive science would likely balk at this strongest conclusion, but there are at least two ameliorative points to be made along the way. First, even if the memory scientist does not leap to the strongest, categorical rejection of inner processes as explanations of remembering phenomena, Wittgenstein offers her good reason to stay on her guard. In any given case she might ask whether *these particular mental processes* are needed, whether we have evidence of them or whether they are just products of immediate theorization. We might treat inner subprocesses as suspicious by default, as unnecessary-and-obstructive until proven otherwise.

Second, it is important to understand the variety of Wittgensteinian rejections on offer. There is a danger that cognitive scientists will characterize the "Wittgensteinian" position as this strongest case: the rejection of any mental processes standing between conditions and manifestations. This is damaging insofar as it keeps any researcher from asking about the necessity and viability of *particular* subprocesses in a *particular* explanation. It is important to see that

Squire's rejection of storage and retrieval for explanations of amnesia, or Matzel and Miller's rejection of carved subprocesses for explanations of amnesia, or any of various researchers' rejections of storage and retrieval for explanations of memory processes generally, are just ways in which contemporary memory research is already in accord with Wittgenstein's advice.

Wittgenstein argues against the need for an inner process in many contexts, but he also argues against the very notion of these inner processes. After all, memory processes like encoding, retrieval, or consolidation are taken to exist at the nebulous cognitive level—not neural, not phenomenal, and not behavioral. It is difficult to get away from positing such processes, remarks Wittgenstein, even though this does not stem from our firsthand familiarity with them.[351] Nonetheless, this impulse to posit mental processes is one philosophers should resist. We may sometimes experience phenomenal processes of memory, but these are not the "underlying" processes of memory—these experienced, inner processes are just other ways that we might remember.

> Memory: "I see us still, sitting at that table".--But have I really the same visual image--or one of those that I had then? Do I also certainly see the table and my friend from the same point of view as then, and so not see myself?---My memory-image is not evidence for that past situation, like a photograph which was taken then and convinces me now that this was how things were then. The memory-image and the memory-words stand on the same level.[352]

Wittgenstein admits that there are sometimes "inner" processes if by this we mean experienced, phenomenal imagery, but these are neither necessary for, nor fundamental to, the process of remembering. These are just another way that we remember. This is not a denial of remembering as a mental process if we construe "mental process" broadly. Remembering with imagery, remembering with words, or remembering with actions are all certainly examples of cognitive

---

351 "But we cannot get away from forming the picture of a mental process. And *not* because we are acquainted with it in our own case!" *Zettel* §565, original emphasis.

352 *Zettel* §650.

processes. The worry is that by asking about the "mental process of memory", we set up the problem in a way that assumes more than is necessary. The cognitive processes of remembering are just these processes already before us.

> Why should I deny that there is a mental process? But "There has just taken place in me the mental process of remembering…." Means nothing more than: "I have just remembered….". To deny the mental process would mean to deny the remembering; to deny that anyone ever remembers anything.[353]

Remembering is a mental process. Indeed, it can be instantiated by all kinds of processes, at least most of which can be rightly described as "mental". But an inner, subconscious process of recollection or retrieval is neither explanatorily helpful nor even clear. After considering all of our ways of remembering, Ayer comes to the same point about the way that positing unconscious processes of memory can be misleading.

> One may say [past events] are recollected unconsciously, if one means no more by this than that one's present ability to remember is causally dependent on them, and so in its way a sign of their having taken place. But it would be much less misleading to say that they are not recollected at all.[354]

Ayer is not alone; Wittgenstein also warns us of the way that this picture can mislead us. What we ought to be doing, according to Wittgenstein, is studying the actual ways that remembering processes occur, and the ways that we use the concept of remembering. Instead, we preemptively decide that remembering must be an inner mental process, and this decision keeps us from apprehending memory for what it is:

> What gives the impression that we want to deny anything? When one says 'Still, an inner process does take place here'--one wants to go on: 'After all, you see it.' And it is this inner process that one means by the word 'remembering'.--The impression that we wanted to deny something arises from our setting our faces against the picture of the 'inner process'. What we deny is that the picture of the inner process gives us the correct idea of the use of the

[353] *Philosophical Investigations* §306.

[354] Ayer 1956, p. 136.

> word 'to remember'. We say that this picture with its ramifications stands in the way of our seeing the use of the word as it is.[355]

Wittgenstein urges us to look at what actually happens when people remember. We may recognize an acquaintance by simply not being surprised when we find her right where we expected to find her. We may recognize an acquaintance after a long, deliberate, internal (and experienced) searching through all the possible people she might be. Most importantly, we might construct limitless cases of recognition that share similarities with either or both of these, ranging from cases where nothing happens at all—and that was recognition—to very calculating inner processes. We err by taking the calculating inner process as a paradigm case, and then by assuming that such a process must simply be *unconscious* when we are not met with it in some case. By this move, argued Wittgenstein, we gain nothing. If we felt dissatisfied by the case where the man simply chose the right bolt of fabric, as though this needed further explanation, no amount of additional processes will alleviate our dissatisfaction. We will, or should, feel just as dissatisfied after the addition of a mental process of remembering where the subject simply produces the right memory trace. Not only are the inner processes of remembering unnecessary, they cannot deliver that result for which they are recruited.

Thus Wittgenstein's remarks on memory are in direct opposition to the computationalist assumptions outlined above. Instead of dividing remembering into distinct systems and componential subprocesses, Wittgenstein's emphasis on the richness in the variety of remembering seems to assure us that this kind of distinct classification will not succeed. Instead of an automatic need for an encoded neural trace that is isomorphic to past events, Wittgenstein's remarks induce us to accept at least the possibility of psychological regularities that are mediated across time

---

[355] *Philosophical Investigations* §305.

without intervening physical regularities. Instead of the assumption that remembering must consist in partly hidden cognitive processes that are implemented by particular neural processes, Wittgenstein's arguments reveal that such hidden processes are unnecessary and ineffective.

4.2 A case study of repertory memory

Before returning to traces and inner processes directly, let us examine one more empirical puzzle about memory. This is not a long-standing impasse, but a very select class of remembering phenomena that dwell near the boundaries of modern memory systems. By virtue of this, these cases expose certain weaknesses in the memory systems framework. The cases in question belong to what is sometimes called repertory memory: the ability of stage actors to remember their lines, and the performance of this ability. Although the memory science literature on repertory is small, these cases are interesting to theoretical memory science for a number of reasons. Repertory is a natural venue for deliberate memorization, with its own elaborate array of culturally developed techniques. As such, it certainly makes a good target for a study in the cognitive ethology of human remembering behaviors.[356] The items remembered in repertory—lines from dialogues, plays, etc.—are some of the most complex memory items to be found. Given that memory scientists are concerned with whether laboratory remembering studies sufficiently complex remembered items, this makes repertory an interesting counterpoint. Most importantly, however, repertory abilities fall very near to the boundary between declarative and nondeclarative remembering.

The distinction between declarative and non-declarative remembering is perhaps the single most popular and trusted distinction between ways of remembering in all of the study of memory,

---

[356] See Intons-Peterson and Smith 1987 for an example.

among both scientists and philosophers.[357] Declarative memory (which, in turn, is usually divided into semantic and episodic memory) is taken to utilize different brain regions, to be implemented by different neural mechanisms, and to be expressed in different behaviors from nondeclarative remembering.[358] There are some scientists—as well as some philosophers—who doubt whether these are even usefully placed under one general category of "remembering".[359] C.D. Broad doubted whether procedural, non-declarative remembering was even cognitive.[360] If these systems happen to occasionally hybridize in complex tasks, the standard view is that this is no more than a contingent overlap of very distinct systems.

There are occasional countercurrents opposed to this consensus. Russell sometimes emphasized the uncomfortable similarities between remembered facts and, in his words, "bodily habit":

> Our habitual knowledge is not always in our minds, but is called up by the appropriate stimuli. If we are asked "what is the capital of France?" we answer "Paris," because of past experience; the past experience is as essential as the present question in the causation of our response. Thus all our habitual knowledge consists of acquired habits, and comes under the head of mnemic phenomena.[361]

Russell concludes this analysis by tentatively preserving the distinction, but not without reservations. As we have already seen above, Ayer drew on Russell's analysis of memory but

---

[357] A distinction that is at least very similar is sometimes drawn between implicit and explicit memory. Whether this is or ought to be the same distinction is an interesting question, but for present purposes I shall keep to the declarative/nondeclarative distinction.

[358] Eichenbaum 1997; Squire 2004; Michaelian 2011b.

[359] Bergson 1896; Broad 1925; Moyal-Sharrock 2009; Michaelian 2011b.

[360] Broad 1925, p. 269.

[361] Russell 1921, p. 80.

ended up concluding that habit memory and declarative memory could not actually be conceptually separated. At least some of the Wittgensteinian arguments above support this same conclusion.

Repertory memory—producing certain sentences that are "called up by the appropriate stimuli"—includes cases that lie just between these. Repertory performance is trained, cued, and perhaps bodily in the same ways that nonverbal skill memory is, and yet the actual performance is verbal. For these reasons alone, repertory is an interesting target, and we will do well to examine this interesting class of borderline cases if only to sharpen our understanding of the distinction. Investigating repertory memory will also present a challenge to the received view about the relationship between embodiment and each of these two ways of remembering. It is unsurprising that there is taken to be an asymmetry here: the acquisition of skills, and especially of straightforwardly bodily skills, clearly depends—deeply and intimately—on the body. The expert athlete or dancer performs a bodily feat as much as a mental one, and the practiced ability to play a piano sonata, or to paint a portrait, depends crucially on training the body in the appropriate ways. Recent data have sometimes demonstrated this bodily dependency to striking extent, but it is safe to say that the general claim that the mastery of skills and procedures requires profound bodily involvement brooks little argument even without such data. Many, if not all, forms of nondeclarative memory are accepted to be embodied, and some of these capacities are in fact taken to be exemplars of embodied cognition.[362]

On the other hand, declarative memory capacities are often taken to be paradigm examples of internal, non-embodied cognition. That is, if you are asked to bring an episode from your past to mind, or to recall the date of your father's birth, whatever process goes on is often assumed to

[362] See Sutton 2007 for recent examples of some of these claims, common in discussions of skill memory and sport, in particular. Sutton's article actually challenges the view that skill memory--even in sport--is entirely bodily.

be purely cognitive, so to speak. While your ability to recall a fact or an episode might be cued by body position or environment, and it certainly may require certain physiological conditions to be met, the process of remembering as such in these cases is not indicative of the embodied agent. The ability to bring facts or experiences "to mind" remains one of the cognitive capacities best suited to a traditional, inner, representational, characterization of cognition. It is true that declarative memory is typically expressed in speech, but researchers' views on the precise nature of the relationship between declarative memory, "inner" cognition, and verbal expression are unsettled and often problematic.

This is hardly a criticism of memory science in particular; devising an unproblematic analysis of the relationship between speech and cognition—as we have already seen above— may be one of the most difficult, contentious, and central problems in cognitive science. Memory scientist Howard Eichenbaum attempts to remain relatively theory-neutral in his definition:

> The defining features of declarative memory...are found in its mode of expression—the ability to bring facts and experiences to mind, that is, to consciously recall items in memory—and then to express the recalled memory in a variety of ways, most prominently by verbal reflection on a learned fact or past experience.[363]

In identifying speech as a way to "express the recalled memory", however, Eichenbaum's definition implicitly commits him to the view that the memory itself is a cognitively internal process or phenomenon. The controversy herein is difficult to avoid, but what matters to memory scientists seeking to study declarative memory capacities is that these are typically only accessible via their expression in language.[364] Speech is the mark of declarative memory, even if "the memory" itself is an inner, disembodied, cognitive process.

---

[363] Eichenbaum 1997, p. 549.

[364] This is only accepted to be typically--or characteristically--the case. In section 3.1.2, we encountered a part of the significant array of work on "declarative" memory processes in nonlinguistic subjects.

However, this neat parallel between the inner/outer distinction and the declarative/nondeclarative distinction blurs, as we investigate borderline cases. In particular, recent data concerning the memory abilities of stage actors provide interesting material at just these junctions. This remembering is thoroughly verbal in nature—it just is the ability to reproduce words and sentences—and yet it shares many key traits with other, nondeclarative, skill memory capacities. All the same, the execution of certain repertory memory abilities can also be very difficult to distinguish in kind from certain standardly declarative examples of remembering—it is not solely by dint of their verbal nature that the actors' repertory abilities might count among declarative memory capacities. Given this dual allegiance repertory memory apparently maintains, a study of its properties will sharpen our understanding of both declarative and nondeclarative memory. As an embodied but verbal skill, this study also gives us pause in too readily circumscribing declarative memory as wholly disembodied.

It has been long known that bodily enactment enhances our ability to remember action-sentences. That is, should you be presented with the sentence 'open the jar' and later asked to recall it, your performance would improve were you to actually open a jar at the time of presentation. Findings from these subject-performed tasks, as they are called, suggest that these gains are independent of mere encoding specificity—the well-documented fact that similarity of states between "encoding" time and "retrieval" time enhances recall abilities—that enacting these items at encoding time strengthens the retrieval capacities regardless of state similarities.[365] In a sense, it would have been surprising had this not turned out to be the case: memory scientists had already known that deep or multimodal processing at encoding improves retrieval, and to enact an action sentence or phrase surely fits into the broad rubric of "deep" processing. Nonetheless, the

[365] See Nilsson 2000 for a review of the literature on subject-performed tasks.

extent of the improvement gained by enaction and the breadth of the cases where such improvement is found has been important, and the demonstration that recall of sentences can, at least in these narrowly circumscribed cases, depend profoundly on the bodily involvement of the subject has already complicated our account of declarative memory.

A recent series of findings primarily found in the work of Helga and Tony Noice, concerning the memorization abilities of actors, has further complicated this account.[366] Not only does the actor's memorization of lines exhibit the same gains with enacted sentences: bodily involvement significantly improves retrieval performance even when there is no literal connection between the remembered line and the action taken. That is, the mere fact of physical involvement facilitates repertory memory. This is true even when the bodily motion is arbitrary,[367] and the gains are even greater when the bodily involvement consists in embodiment of the role—again, even when this embodiment does not have the literal connection to the remembered line that is found in the subject-performed task experiments.[368] Embodying the role physically enables the actor to remember the script, and this physical embodiment is already a part of the learning methods employed by actors. Given the nature of the trained skill of learning scripts, it again would probably have been more surprising had this turned out not to be the case. Noice and Noice conclude that the memory processes of actors are just instances of the kind of embodied cognition usually associated with nonverbal skills.

If the mark of declarative memory is verbal expression, and the process of declarative memory is supposed to be the inner, non-embodied retrieval of a mental item, then the embodied

---

[366] Particularly Noice and Noice 2006, but also, for example, Noice et al. 2000.

[367] Noice et al. 2000.

[368] Noice and Noice 2006.

verbal memory skills revealed by these data already provide a memory process ill-suited to current classifications. What these actors reproduce is thoroughly verbal, and yet the learning and performance processes appear to be deeply embodied skills. If these memory performances are admitted into the fold of declarative memory, then a cognitively internal, disembodied characterization of declarative memory will be untenable. The theorist who seeks to preserve the domain of declarative memory from embodiment must avoid the conjunction apparently offered by these data. There are such responses available, but none is without consequence.

The first response, and probably the one that has the most appeal when first confronted with the cases of repertory memory, is to simply reclassify repertory memory as unproblematically nondeclarative. Although the performance of this skill involves words, it is not a memory that is expressed in words in the same way as standard declarative memory, our defender of disembodied declarative remembering might claim. The skill that the actors are mastering only employs sentences incidentally, as it were, and not as a report of an inner, mental remembering process. While sentences happen to be produced as an artifact of the particular skill, the process in which the actors engage is properly classified with learning to produce particular sounds on cue—repertory memory should hence be typed with the implicit memory active in trained musical skills. The surface characteristics may have briefly tricked us into thinking that repertory is essentially declarative, but reflection on its properties—even on the very properties here emphasized—reveal its basically nondeclarative nature.

There is much to be said for this response. There is indeed the danger of a kind of equivocation when treating the remembered lines actors produce as factive. That is, the actor only remembers that the next line is 'We are such stuff as dreams are made on' in the same way that the pianist remembers that the next chord is an F-sharp-minor-seventh. We can, in a certain contrived fashion, describe these as "remembering that", but when the performance comes off, we only mean by this to claim that the subject has remembered how to go on. It is notable that both actors and musicians report a marked absence of inner, mental, what-comes-next kinds of steps in the process once they are proficient. This is a point Hubert Dreyfus has discussed at length in his work on human learning and expertise.[369] In this respect, there is certainly a tighter relationship between repertory memory and particular instances of nondeclarative skill memory than there is between the actors' abilities and garden-variety semantic or episodic memory. To group repertory memory with declarative abilities merely as a consequence of the fact that sentences are produced would be to read too deeply into this superficial trait.

However, the mere fact that sentences are produced is not the only tie between repertory and standard declarative memory. At least at first glance, there appears to be a deep similarity between the trained verbal responses of the veteran actor and our own verbal responses to common conversational questions. That is, although remembering the date of one's birthday is taken to be a paradigmatically declarative (semantic memory) ability, the process of actually responding to the question "What's your birthday?" looks very like the trained actor responding to his cue. The fact that the theorist who responds with this first rejoinder now has to be careful to define declarative memory so as to exclude repertory is already significant. The fact that she has to define it so that it excludes repertory responses but includes birthday responses may be insurmountable.

There is a second response. One may admit repertory as a declarative memory phenomenon but to make an exception of it--to deny that the embodiment we find in these cases is particularly problematic or indicative of declarative remembering more generally. It is a special case of declarative memory, the defending theorist might claim, given that the sentences uttered

---

[369] For example, see Dreyfus and Dreyfus 1980, but this is also a key theme in Dreyfus 1972/1992. The same point, about memory in particular, is made by Ayer in the *Problem of Knowledge* (1956, p. 135).

were trained in so skill-like a fashion. While technically declarative, we ought to just accept this—admittedly very narrowly circumscribed—case as that one in which declarative memory is embodied.

It is difficult to keep this response from being too *ad hoc*, as it is. Something like this might even turn out to be the right way to go about classifying these types of memory, but until we have grounds for making exceptions of particular cases it will be hard to make detailed distinctions like this without the lines appearing to be arbitrary responses to the problem that has been raised. Wittgenstein warns us not to hold a "contemptuous attitude towards the particular case" (*Blue and Brown Books*, p. 18), and sweeping the idiosyncrasies of boundary cases under our theoretical rug is having just this attitude. Furthermore, embodiment is not the only aberrant trait in repertory memory; certain aspects of stage recall genuinely seem to fit the skill memory rubric better than they fit into our understanding of declarative memory.

These considerations lead us to another formulation of the motives behind this second response: perhaps it is best, our defender of disembodied declarative remembering might say, just to admit that we have hit upon a problematic borderline case and leave it at that. There already exists much discussion of just how deep we want to claim these memory system divisions run,[370] and perhaps we can maintain much of what is beneficial in the declarative/nondeclarative classification system even in the face of a few isolated and ill-behaved borderline cases. So memory phenomena have been identified that exhibit both declarative and non-declarative traits? So be it. Nobody had claimed that declarative memory conceptually required disembodiment, but only that such was typical of these memory processes.

---

[370]See, for example, Michaelian 2011.

There is something to be said for this response as well. As Dr. Johnson is reputed to have said, the fact of twilight does not negate the dawn. That is, the existence of ambiguity at the boundaries does not mean that there are no clear cases—it does not even mean that most cases are not clear. Problematically borderline cases can tell us a lot: if one wants to know how night is related to day, or just what a dawn consists in, it is exactly the borderline, twilight, cases to which one ought to direct one's attention. A lucid study of borderline cases can sharpen our understanding of both of the concepts in question as well as all the instances that fall under them, and it can be crucial to better conceptual organization to struggle to analyze these cases. Nonetheless, there are sometimes simply ill-behaved boundary cases; it might be that repertory memory is some kind of thoroughly hybrid activity, or it might be that the divisions we have carved out of memory break down at points like these.

This response does deserve careful consideration, but in the end it falls prey to some of the same consideration that gave us pause in attempting the first response. First, we might note that the inclination to force ill-behaved particular cases into the category to which we "already know" they belong is just the predisposition that Malcolm, and Wittgenstein, warned about in the last section. Wittgenstein's point is well taken that we should pause, at least, before behaving so contemptuously toward the particular case. Furthermore, if exception is made on the basis of the borderline status of repertory memory, our questions about disembodiment will just advance to the next nearest ability in the direction of the core of declarative memory. Once more, the learned responses so common in our everyday conversations appear to share many of the same characteristics. In fact, it is hard to see how our response to a question about our name or birthday could be cognitively distinguishable from the trained responses of actors who have learned a script. This is not to say that there are never phenomena that are so distinguishable, of course: there are instances in which recalling a fact does seem very different from the trained, effortless, skilled responses that expert actors (and expert speakers) demonstrate. All the same, at least some of the standardly accepted cases of declarative memory appear to exhibit these nondeclarative traits, regardless of the exceptions made for actors. To some extent, of course, this last is an empirical claim. To probe the role of embodiment in our declarative responses to everyday questions, however, is no straightforward task. Given that experiments involving the performance of successfully learning conversational responses and the role of bodily involvement therein are probably practically implausible, this may only be in principle susceptible to empirical confirmation. Even if this remains undecided, though, the analysis of these possible responses has left us with a few interesting upshots.

There is the basic worry that repertory memory is declarative yet embodied. It is worth noting that even if either of the above responses to this worry succeeds utterly, progress has been made. If we accept the first response, we are charged with formulating a definition of declarative memory that retains what we seek to retain and excludes what we seek to exclude. As mentioned above, this is no easy task, but it may be feasible. If we can characterize declarative memory in terms that include responding with one's birthday but that do not include responding with one's line, we will have elucidated the concept considerably. Similarly, if we are able to provide grounds for the special status of repertory memory—which may come down to analyzing its hybrid nature or being able to rigorously draw out some of the underlying framework for declarative and nondeclarative capacities—we will have come to a clearer understanding of the relationship between these two ways of remembering. Furthermore, if sufficient explanation can be given for why conversational responses do not deserve this special multiform or hybrid status, we will have done a service to our understanding of the role of speech in declarative memory.

Given some of the data and considerations above, however, there is reason to remain pessimistic concerning this last possibility. The unifying theme in revealing the difficulties for any of these above responses to the basic problem is the similarity between repertory memory and conversational responses. There was never any doubt that language is a skill—albeit a uniquely intellectual skill—and the fact that linguistic memory performances, of all kinds, inherit skill memory characteristics is, in some ways, unsurprising. There is already work in other areas of language and memory that emphasizes this conjunction: learning artificial grammars also follows nondeclarative, skill memory patterns.[371] Recent findings in neuroscience have, notably, indicated the same brain area activations for both language and memory.[372] If Wittgenstein and Ayer are right that sometimes when we say the date of our birthday in response to a query, the speaking of it *just is* the remembering in the same way that the pianist playing the next chord just is remembering the next chord, any attempt to separate the verbal "report" from the trained skill will be strictly blocked.

Embodiment aside, the whole array of repertory and repertory-like cases demonstrates profound difficulties with the way that memory systems are standardly treated. To train someone to respond to given cues with certain bodily actions is taken to be a paradigm case of non-declarative, procedural remembering. To verbally recount learned information is taken to be a paradigm case of declarative, semantic remembering. The above cases show that there is no easy way to draw a line distinguishing these. More importantly, any attempted boundary would only give rise to further intermediate and hybrid cases. Wittgenstein asks us to consider someone who is asked whether they have ever been in this room before. The subject responds, verbally, without

---

[371] Knowlton et al. 1992.

[372] Gabrieli et al. 1998.

any calculation or thought. Without an internal process of bringing a past episode to mind, this will be indistinguishable in kind from repertory. This is not because it is a special case, however, it is just that remembering may be comprised of the most various occurrences. If we are inclined to expand the boundary of non-declarative skill memory to include repertory and basic trained verbal responses, once more we can isolate intermediate cases. The remembering subject might get the answer right and then go on to explain his previous time in the room in question, just as Wittgenstein details. Once again the boundary destabilizes. There are cases that cannot be said to be merely non-declarative or merely declarative, and accepting this indeterminacy is just what Michel ter Hark suggests that Wittgenstein is urging us to do. Psychological indeterminacy does not—at least does not always—stem from there being a hidden and unknown component to the process in question. Rather, psychological processes are essentially and importantly indeterminate in this fashion.

4.3 Extirpating Componentiality, Traces, and Hidden Processes

The computationalist assumptions introduced in chapter two are at the heart of each of the philosophical and empirical difficulties above. In the following, let us examine each in turn and the role that it plays in causing conceptual and experimental difficulties for the study of memory. Implicit reliance on computationalism is impeding progress and clarity in the cognitive sciences of memory. For each of these assumptions, it will be demonstrated that there is no cause to automatically accept the claim, and furthermore that there are good reasons to reject it.

4.3.1 *Horizontal componentiality*

As we have seen, the insight that remembering may consist of many different activities, states, and capacities is one Wittgensteinian lesson about the careful characterization of memory and many other mental phenomena, and this lesson has, in some sense, already been heeded. This is not to say that this insight originated *ex nihilo* with Wittgenstein; as is noted above, many philosophers before and after Wittgenstein—some of whom share little else with Wittgenstein—have remarked on the rich varieties of remembering. Psychologists and other scientists of memory have explicitly drawn on these ideas when approaching the study of memory, and have hence divided our ways of remembering into particular memory systems. We would be mistaken to study the system of memory as if it were just one thing, and so the study of memory must study several distinct systems.

This last move, however directly and often it has been made, does not follow without the computationalist assumption of horizontal componentiality. It is true that there are many systems that can be rightly said to engage in processes that are decomposable in this manner. Sitting here at the computer, I can save my work to the computer's hard drive, or I can save it to a shared drive on the network. These are both instances of "saving work", but each is served by a software and hardware process that is distinct from those processes serving the other. If we want to understand how work is saved on a computer, it will do us well to first understand that work can be saved in a few different ways, and then to investigate each of these systems in turn. Similarly, the first lesson in understanding how cells can divide is understanding that there are two distinct processes—mitosis and meiosis—that underlie cell division. Although there are similarities at a coarse-grained level of description, cell division cannot be understood until each of these component processes is independently understood. In the face of examples like these, it might seem as though any phenomenon we are seeking to explain must either be unitary and explained as a single system, or componential and explained as multiple systems.

However, it is clearly not the case that all non-unitary mental phenomena require component systems. To see this, we might compare remembering to certain other mental phenomena. One case to which Wittgenstein also devotes his attention is *expecting*. In the *Philosophical Investigations*, Wittgenstein repeatedly emphasizes and considers the different ways in which we might rightly ascribe "expecting" to someone.[373] My expectation of someone or something may consist in a sustained feeling of eager anticipation, but it may alternatively consist in any number of other expressions—talking about the expected person or thing, looking in the direction from which this expected thing is anticipated to arrive, or even just the absence of surprise when the expected event occurs. There are different ways to expect, but it hardly follows that there are distinct "expectation systems."

Wittgenstein is right to caution us against forgetting or ignoring that these phenomena are so multiply manifest, both in the case of expecting and in the case of remembering, but note that in the case of expectation there is no immediate inclination toward component systems. Cases of expecting are not all alike, but this does not mean that there are distinct, independent processes of expectation, driven by distinct underlying processes. Somebody who took Wittgenstein's lesson to heart by deciding that there must therefore be, for example, a system of declarative expectation and a system of nondeclarative expectation, would have absurdly misconstrued the implications of the varieties of expectation.[374] Furthermore, we might construct very similar examples out of

---

[373] See *Philosophical Investigations* §§577-586.

[374] Anna Boncompagni makes a similar point about Wittgenstein's discussions of aspect-seeing and their application in debates concerning the conceptual vs. nonconceptual basic character of cognition. To take Wittgenstein's examples as straightforward examples of distinct categories is to ignore his emphasis on comparison and differentiation among a spectrum of interrelated cases (Boncompagni 2013, p. 44).

any number of psychological verbs: hoping, believing, desiring, etc. In all of these cases it can be helpful, even in psychology, to underscore the wide variety of expression by which these phenomena manifest. However, in none of these cases is it naturally assumed that the phenomenon's variety of forms reveals the component systems or processes that constitute it—nor would it be helpful to assume as much. In cases like cell division or computer processes, it is naturally and helpfully assumed that variety of forms reveals underlying distinctions among component systems and processes. So one question for those who study memory is simply whether there is, *prima facie*, better reason to count remembering among the componential-friendly phenomena like cell division or among the componential-unfriendly phenomena like expectation. At the very least, it is not obviously the case that memory is more like cell division than it is like expectation.

To be sure, the fact that two of these are cognitive or mental phenomena is probably not sufficient to conclude that an explanation of memory should not assume horizontal componentiality. There may be examples of mental phenomena that lend themselves to horizontally componential explanation. Feeling pain, for example, seems to happen in at least two distinct physiological ways, nociceptive pain and neuropathic pain.[375] Although the details are still not complete, it appears that these are served by relatively distinct physiological systems and processes, and that the study of pain has benefited by the discovery of this distinction. Pain is not alone; there are other phenomena that are rightly counted as mental or cognitive that really are composed of component, alternative processes.[376] Those who seek to divide remembering into

---

375 A brief survey of the distinction can be found in Charles Vecht's *Nociceptive Nerve Pain and Neuropathic Pain* (1989).

376 E.g., ringing in the ears, or tinnitus, can be manifest by what are at least to some extent distinct, underlying physiological processes.

conceptually distinct component processes are just committed to the claim that memory is more like pain, at least in this regard, than it is like expectation or hope. However, there is no immediate reason why memory should belong among the latter rather than the former.

This admission alone, that it is not yet clear whether horizontal componentiality is a suitable framework for understanding memory, already constitutes significant progress. Rather than assuming without argument that different ways of remembering must mean distinct memory process types, we must now actually look at the phenomena of remembering, and at how this hypothesis of componentiality fares in the face of empirical and conceptual considerations. Although I will argue that it has fared poorly, and that we have good reason to count remembering among those phenomena that do not exhibit robust horizontal componentiality, the mere exposure of this underlying assumption to doubt and scrutiny can only help to clarify and to resolve those difficulties surrounding memory systems.

Although the division of memory systems has been one of the developments in memory science that has brooked less disagreement than most, we can see from the considerations above that it is not without difficulties. Most cognitive psychologists studying memory report that distinct memory systems are served by distinct brain systems, but cognitive neuroscience has only turned up patchy and sometimes contradictory evidence for this.[377] What it is that makes systems distinct is taken in many cases to be the different neural mechanisms that implement each system, but as we have already seen, the neural mechanisms implementing these have yet to be discovered. Additionally, it looks as though there are inexorably intermediate cases plaguing the boundaries between systems.[378] Given all of these, we should not only to reject horizontal componentiality

[377] Kim and Baxter 2001; Johnson 2007.
[378] These are also briefly characterized by Hacker and Bennett (2003, p. 155).

as a automatically warranted assumption in the study of memory, but we should also grant memory prima facie membership among the class of cognitive capacities—like expectation or hope—that we do not expect to be able to decompose into systems.

4.3.2 *Vertical componentiality*

In addition to the assumption that alternative and distinct component process types are on offer for any given instance of remembering, we have also seen that researchers tend to characterize what processes there are as vertically componential. The explanation of the memory process is taken to require an explanation of those subprocesses that form the phases of remembering. Once again, this strategy has been seldom questioned. There has been much discussion of the best characterization of these subprocesses, the ways in which these subprocesses are supported or composed neurally, and of how these subprocesses function and fail to function in cases of remembering and forgetting. There has been little discussion of the assumption that some chain of subprocesses must explain memory.

The considerations that destabilize the vertical componentiality assumption run parallel to the considerations against horizontal componentiality in remembering. There are processes that are amenable to vertical decomposition and there are processes that are not so amenable. These categories are not coextensive with the groups of processes that are or are not amenable to horizontal componentiality: stalemate in chess, for example, can be accomplished by either leaving an opponent with no legal moves or by taking the last non-king piece. These are alternative ways of stalemating, and as such we might rightly say that stalemating is horizontally decomposable. Nonetheless, stalemating in chess does not seem to be amenable to *vertical* decomposition. That is, there are not meaningful *stages* of either of these processes. There are precursors, but these are not stages of the stalemating process. The move that begets stalemate does take time, and physically moving the piece from one square to another progresses through a series of states, but there is no interesting sense in which these states are *components of the stalemate*. The stalemating just happens, as it were, either completely or not at all.

Once more, the mere existence of processes like this should be enough to call into question the assumption that remembering processes *must be* vertically decomposable. Some processes decompose in this manner and some do not. Again this is not a matter of whether the process in question is cognitive or not. Understanding, for example, is no more amenable to stages than is stalemating, at least in many cases—neither of which are non-cognitive abilities. In a note in the *Philosophical Investigations*, Wittgenstein illustrates this difference by emphasizing the role of duration in the respective psychological ascriptions.

> (a) "Understanding a word": a state. But a mental state? — We call dejection, excitement, pain, mental states. Carry out a grammatical investigation as follows: we say
> "He felt dejected the whole day"
> "He was in great excitement the whole day"
> "He has been in pain uninterruptedly since yesterday". –
> We also say, "Since yesterday I have understood this word." 'Uninterruptedly', though? – To be sure, one can speak of an interruption of understanding. But in what cases? Compare: "When did your pains get less?" and "When did you stop understanding that word?"
>
> (b) What if one asked: When can you play chess? All the time? Or just while you are making a move? And the whole of chess during each move? -- And how odd that being able to play chess should take such a short time, and a game so much longer![379]

Without sufficient scrutiny, it can seem like we might be dejected by something or we might understand something, and that these are two alternative states or processes that may go on inside our mind. What Wittgenstein seeks to show us is that this similarity is merely a surface feature of a comparison of the two psychological ascriptions. Despite what we might first think, understanding is not a state or process that we undergo. The most varied things can constitute

---

[379] This is the complete text from an aside between §149 and §150 in the *Philosophical Investigations*. The context is largely a discussion of the nature of mental states, mental processes, and understanding.

understanding—Wittgenstein's surrounding remarks make it clear that simply *being able to go on* is an important example. One way to get a fix on the difference is to inquire about duration. Asking how long one has spent understanding something reveals that understanding cannot be the sort of thing that pain is. Some cognitive phenomena are processes or states with particular durations, and some are not. Those which are not will not easily lend themselves to stages or subprocesses.

It is just so for remembering. We may spend a very particular amount of time trying to remember something we cannot remember. We may spend a given duration reminiscing about some past episode. We may spend a certain amount of time attending to inner visual imagery that represents a past scene. In any of these cases, we can loosely be said to "be remembering" for a certain amount of time. However, if we remember how to play chess, or if we raise our hand because we remember the answer to a question, or if we can rightly be said to remember our own tenth birthday, but not our fourth, we do not engage in any process or state whose duration can sensibly be asked about. This does not simply mean that some remembering has duration and some does not: the former cases are all cases where other processes or states are bearing the duration. We spend time trying to remember—time rehearsing or speaking or imagining.[380] We spend time attending to imagery, which also happens to instantiate remembering in the same way that moving the pawn—which takes time and has stages—instantiates stalemating—which does not. When we spend time reminiscing about the past, we are actually engaged in attending, imagery, speaking or inner rehearsing, and emotional responses, all of which do have measurable and sensible durations. We might remember with imagery, or with speech, or with a raised hand.

[380] It is not uncommon for cognitive scientists to study the properties of *trying* to remember a stimulus, and then to draw conclusions about the nature of remembering. For example, the influential article "The Cognitive Neuroscience of Remembering" by Randy Buckner and Mark Wheeler leans heavily on methodologies that measure activation during the conscious process of *attempting* to remember difficult stimuli (Buckner and Wheeler 2001).

Each of these activities in the world takes time and is accompanied by certain behaviors or phenomenologies, or both, but none of them is *what it is to remember* more generally.

The realization that statements about remembering's duration more closely track the structure of psychological ascriptions like understanding, belief, or expectation than those durational processes or states like pain or intense emotions more than destabilizes the assumption that memory must be vertically decomposable into constituent subprocesses. A process that cannot be said to have any precise duration does not lend itself to being carved into stages—the very idea of stages includes a notion of succession, and without duration there is no room for this. Any particular instance of a non-durational psychological phenomenon may be *instantiated* by other bodily or phenomenological processes that can have measurable durations and phases, but this does not mean that we can identify stages or subprocesses characteristic of the phenomenon-*type* itself. We may remember with words, or with images, or with actions. That these have durations and component phases does not imply that *remembering* also has these. We might also, for example, *respond* with words, or images, or actions—but to take this as warrant to study the characteristic stages of "responding" is simply to misunderstand what we mean when we ascribe "responding".

This is not to say that subcomponents and duration have no place in the study of memory, of course. There may be interestingly robust groups of memory phenomena that are implemented or instantiated in similar enough ways that schematic phases can be described. Repertory memory, for example, can be studied in this way. Veteran stage actors have an established body of techniques for memorizing, and these have natural stages. Trying to learn or remember something by consciously using specific techniques will readily lend itself to duration and stages, but these stages are not the stages of memory more generally. Remembering is made up of a wide and dense variety of phenomena, many of which are non-durational and non-compositional.

4.3.3 *A memory trace: the very idea*

Let us revisit Tulving's definition of the memory trace, which has been influential, in part, due to its admirable lack of pre-empirical empirical content. Tulving defines the trace as that neural change that "accompanies a mental experience…whose retention, modified or otherwise, allows the individual later to have mental experiences of the kind that would not have been possible in the absence of the trace."[381] What Tulving and those who accept his definition are trying to accomplish here is to allow for use of the concept of a memory trace that plays a key role in the remembering process, but without committing themselves ahead of time to any particular *form*—neural or structural—this trace must take. This is a worthy goal, and one that stems from, and contributes to, many of the considerations above. I will nonetheless argue that, as framed, this goal cannot be satisfactorily attained.

Tulving defines the trace as that neural change accompanying a mental experience that is a necessary physical condition for later mental experiences.[382] These later "mental experiences" must, of course, be merely dispositional rather than actual. That is, whether a neural change counts as a memory trace or not cannot depend on whether it is *actually* activated later. If this were so, the "inactive trace" would be a logical impossibility, since nothing could be (or could have been) a trace until it were actually activated. Strictly speaking, Tulving's definition should invoke the subjunctive mood a bit earlier, and define the trace as that neural change that *would allow* the

---

[381] Tulving 2007, p. 66.

[382] Tulving's phrase 'mental experience' is itself rife with its own points of philosophical departure, but for the present purposes let us grant the sensibility of this locution.

individual certain mental experiences later. Tulving and others, however, already seem to treat the trace as this: that neural change that would allow, under various conditions, various later mental events. This technicality, however, is not trivial. Explicitly noting the dispositional nature of the trace will reveal a flaw in the underlying strategy at work.

The only condition on the later mental event in Tulving's definition is that it requires the neural change that accompanied the earlier event. This is also a deliberate austerity: since the trace is a key part in giving an account of memory, the memory scientist seeks to avoid invoking the notion of memory in the definition of a trace—to do otherwise is dangerously near to vicious circularity. However, this means that in order for a neural change to count as a memory trace it must fulfill just two conditions:

(1) The neural change must accompany a mental experience.

(2) There must be some conditions under which a later mental experience would occur that could not have occurred without the neural change in question.

But what, then, *does not count* as a memory trace? A neural change that fails (1) does not count, and this seems intuitively acceptable.[383] However, any neural change that meets (1) need only also meet (2) in order to count as a trace. What is it like to fail (2)? Only when some neural change is such that *there are no conditions under which any later mental experience would require it* will that neural change fail this second condition. This is implausibly restrictive. The collection of neural changes that accompany mental experience but which could never, under any conditions, have any effect on later mental events is vanishingly small. If the study of the complexity of human behavior and of the profound plasticity of the human mind and brain has taught us anything,

[383] There are probably interesting borderline cases, however we define "mental experience", but let us presently concede that, by and large, those neural changes that are in no way accompanied by mental experiences (perhaps, say, because the subject is asleep or otherwise totally unaware) do not make good candidates for traces.

it has taught us that there no experiences that would not leave detectable ripples under some conditions. If we are to accept Tulving's definition of the trace, then every neural change is a memory trace.[384]

This is especially so given that the instance of 'necessary condition' in this definition has no additional requirement that it be proximal. Any neural change that could be studied by a neuroscientist and reported to the patient in question is just a necessary condition for the possible mental experience had by the patient during that conversation with her doctor. So the only neural changes that will not meet (2) are those that could never be detected in any way. In effect, any neural change that accompanies any mental experience, and can be detected *in any way* will, under this definition, be a memory trace.

Even if this last consideration is dealt with by building some kind of "appropriate" causal connection into the definition, neural changes that meet the resulting conditions will still be disastrously ubiquitous. Given the dispositional nature of (2), it is not clear how any neural change that is even remotely interesting could fail this condition. Every neural change, or at least nearly every neural change, has the *potential* to affect later mental events given the right conditions. So if this is what a memory trace is, those who are looking for traces in the brain need look no further. Indeed, they cannot point to anything that is *not* what they are seeking.[385]

Tulving is a careful thinker, and he deliberately constructed this definition free of any claims of structural isomorphism that would mark the trace out from whatever other neural configurations or changes it might be among. There are good reasons to avoid the commitment to

---

[384] Tulving makes it clear that he does not aim to have this definition capture all of the lasting effects of past experience, but he admits that understanding what could separate memory from non-memory neural activity "remains a challenging problem" (Tulving 2007, p. 66).

[385] Malcolm makes a related point, insisting that for something to be a memory *trace* it must have a tighter relationship to both past and future processes than mere physical continuity (Malcolm 1977, p. 174).

this kind of neural correlate, some of which we have already seen and some of which will be elaborated below. Nonetheless, in the absence of a structural relationship with that which it is supposed to represent, the memory trace can only be characterized such that every neural change counts as a trace. Worse than this, the history of a neural change is very unlikely to suffice for individuating "the change" later in time. If an eight ball is lightly struck by the cue ball and moves across the table in a line, the eight ball's change in position can indeed provide a record, so to speak, of the prior collision. With a few reasonable physical assumptions and the appropriate laws in hand, we can work backward form the change in position to the velocity and momentum of the cue ball at the prior impact. However, as the game progresses, many more impacts occur. If four or five turns later the eight ball has been struck time again, any sense that a present physical change provides a record of the *initial* impact with the cue ball has completely disappeared. It is true that the eight ball would not have its present position were it not for that much earlier impact, and that its present position is necessary for its future behavior, but there is no way to individuate "the change" in the eight ball's position that is due to that original collision. This could only mean something like the difference between the eight ball's actual position and the position it would be in if, other than that original impact, *everything about the game had been the same*. But of course, this is just to talk about a different game. Worse, it is not clear even in principle what this difference could refer to. The more interactions that have intervened since a past event, the less sense there is individuating "a change", at all, that is due to that past event.[386]

These are just reasons why any successful concept of the memory trace must include something very like structural isomorphism. If there does not persist some way to "match" the neural change to the past event—and history alone cannot provide this—then any talk of this neural

---

[386] This is so even for the relatively simple and modular system of a pool table. The profound interconnectivity of the brain and the real features of neural network modification only make for a stronger case.

change as a trace of some event can only offer unfulfillable promises. Perhaps this point is implicitly appreciated by the memory science community, since the trace is often treated as isomorphic even if this isomorphism is not always made explicit in what a trace is taken to be.[387] The isomorphic trace, however, is just the source of the difficulties above: in section 3.3.2, the standstill in empirical investigation into traces or engrams was illustrated. After decades of research—and much, much progress in understanding the brain as well as the behaviors of remembering—memory science seems no closer to identifying traces, the mechanisms of traces, or even the types of rules or translation schemes that would allow us to decode such traces were they found. In section 4.3.2 above, the Wittgensteinian argument for the inefficacy (and superfluity) of such a trace was illustrated. A trace requires isomorphism, but an isomorphic trace is empirically and philosophically problematic. In his *Remarks on the Philosophy of Psychology*, Wittgenstein compares human memory to the records and traces in a machine:

> An event leaves a trace in the memory: one sometimes imagines this as if it consisted in the event's having left a trace, an impression, a consequence, in the nervous system. As if one could say: even the nerves have a memory. But then when someone remembered an event, he would have to infer it from this impression, this trace. Whatever the event does leave behind in the organism, it isn't the memory. The organism compared with a Dictaphone spool; the impression, the trace, is the alteration in the spool that the voice leaves behind. Can one say that the Dictaphone (or the spool) is remembering what was spoken all over again, when it reproduces what it took?[388]

This comparison is doubly interesting today, since the language of "memory" has been extended to cover phenomena similar to this in computer science. Extending 'remembering' to Wittgenstein's Dictaphone seems obviously strained, much as it did when the term 'memory' was first applied to computing. Even after the extension of 'memory' to computer information storage,

---

[387] That the trace is taken to be *encoded* is an indicator of this assumed isomorphism. Hutto and Myin warn of the "high cost" of taking the code metaphor, which is "rife in the cognitive sciences", seriously (Hutto and Myin 2013, p. 70).

[388] *Remarks on the Philosophy of Psychology*, *vol. 1*, §220.

we still do not characterize the computer itself as *remembering* anything. Our own remembering is unlike the recorded information of the Dictaphone or the computer—encoded information about a past event can only contribute to the replication of the past event's structure if this structure is "read off" from the encoded item. But as Wittgenstein's remark suggests, it is implausible to claim that *we* read off the past from these encoded traces. We do nothing of the sort. Rather, it is somehow thought that there is a "subpersonal" decoding at retrieval, which transforms (somehow) into our conscious, personal, remembering behavior.

Once more, demonstrating that the encoded trace is no foregone conclusion already constitutes progress. It is all too often assumed that there *must be* such a trace, and the considerations above show that such a trace is unnecessary. Even if nothing but this destabilization is accomplished, some of the difficulties surrounding the mechanisms and neurobiology of memory may be clarified. We have seen more than this. We have also met arguments for the ineffectiveness of the trace. In addition, nothing in the phenomenology, behavior, or neuroscience of memory has provided evidence for such a trace. Lastly, understanding and investigating the trace has proven fiendishly difficult. I submit that given all of these, the burden of proof lies with the trace advocate to show that there is or need be such a problematic element in our understanding of memory phenomena.

4.3.4 *Mental software*

The assumption that memory is based on, or is supported by, or consists in, inner processes is itself compound, and it can be ambiguous. As we have seen, these inner processes are sometimes taken to be neural, and sometimes taken to be "cognitive". Let us focus on the non-neural processes of remembering first. As we have seen, these processes are also not experienced, phenomenal processes. Rather, they are purported to exist at an intermediate level, implemented by neural processes and in turn causal of phenomenal and behavioral processes.[389] As is the case for many cognitive phenomena, explanations of remembering in terms of inner, cognitive processes must be carefully constructed such that they do not commit homunculus fallacies or invite related regress problems. Many researchers, in all of cognitive science, are well aware of this worry.[390] Sometimes this homunculus takes explicit form as the "central executive" which controls memory systems and processes: Alan Baddeley writes that, "working memory is assumed to be controlled by an attentionally limited component, the central executive."[391] Baddeley admits that this is "somewhat homunculus like", but he assures the reader that this homunculus can, in due course, "be replaced by a set of clearly specified processes."

Wittgenstein worried that the kind of coded memory trace we might tend to believe must accompany remembering could only be useful if we read off this trace somehow in our remembering. Even if this "reading off" is somehow taken care of by a subpersonal "specified process", such a process must be initiated by our present environment and stimuli. But how do the present stimuli suffice to specify the evocation of the trace? As soon as this is explained, we may as well simply explain that the remembering phenomena were so cued by the stimuli. Either we need to stipulate a mechanism by which such memory-reactions are performed, in which case

[389] One salient example of this can be found in Stephen Kosslyn's influential accounts of perception and working memory (Kosslyn and Sussman 1995, Thompson et al. 2001, Kosslyn 2005). Kosslyn and Amy Sussman, explicitly borrowing from David Lowe's work with computer vision (Lowe 1987), solve the "problem" of how we recognize what we see by suggesting that models of objects we have seen are stored in "visual memory". Then, "when activated, the model generates an image"—presumably unconsciously—which is compared to our perceptual input. When the match is sufficiently close, we recognize the object. Kosslyn and Sussman conclude that this generation of images from memory must be "an essential part of our ability to recognize objects" (Kosslyn and Sussman 1995, p. 1036). Their account runs strikingly parallel to the parable of the grocer in the opening of the *Philosophical Investigations*.

[390] In fact, I will venture to claim that popular awareness of homuncular fallacies is one of the most straightforward and tractable success stories among the ways that philosophers of mind and cognitive science have contributed to better science, even if the treatment of homunculi by scientists does not always conform to philosophers' standards.

[391] Baddeley 2007, p. 152.

we will end with an infinite chain of selected traces and pre-traces, or we need not stipulate such a mechanism, in which case the trace is superfluous in the first place. Several philosophers have given versions of this Wittgensteinian argument against the notion of an inner memory process.[392] A very similar argument was given by Merleau-Ponty as well, in his *Phenomenology of Perception*.

> Thus the appeal to memory presupposes what it is supposed to explain; the patterning of data, the imposition of meaning on a chaos of sense-data. No sooner is the recollection of memories made possible than it becomes superfluous, since the work it is being asked to do is already done.[393]

Some memory scientists agree. Gianfranco Dalla Barba calls this problem the "memory-trace paradox", and claims that most mainstream accounts of memory fall into it because of their implicit commitment to a kind of "homunculus fallacy".[394] Once remembering is sketched in terms of a process of storage and retrieval, it must be the case that something is being retrieved. It is not the remembering subject retrieving this thing (for that is the explanandum), and so the retrieval of the retrieved item must be committed subpersonally. Thus, Dalla Barba's homunculus fallacy, leading to the memory-trace paradox, is already launched.

It is not hard to find fairly explicit examples of this among influential work in memory science. Yadin Dudai admits that memory traces must be stored along with "retrieval handles" in order to be activated in relevant circumstances and fashions. Without these handles, the stored item could never be relevantly identified in response to given stimuli. Dudai goes on to admit that

---

392 E.g. Malcolm 1977, p. 93-97; Shanon 1998; Bennett and Hacker 2003, p. 164; Moyal-Sharrock 2009.

393 Merleau-Ponty 1945/1996, p. 23. Again in the next paragraph he writes, "How could the evocation of memories come about unless guided by the look of the strictly visible data, and if it is thus guided, what use is it then…?" Cf. also Malcolm 1977, p. 101: "The memory theorist makes a useless movement. He invents a memory-process to fill what he thinks is an explanatory gap. He is deceived in thinking that some progress in explanation has been achieved. A gap offended him; a gap remains."

394 Dalla Barba 2001.

understanding or modeling these "retrieval handles is mostly a *terra incognita.*"[395] Indeed, one suspects that an effective model of how such handles could operate will continue to be *incognita* for some time, given that the task of understanding how a handle can be retrieved by the system will be no simpler than understanding how a trace could be retrieved by the system. Endel Tulving recognizes this difficulty obliquely, writing, "to assume that the rememberer knows what it is that he wants is to beg the question about retrieval: if the nature of the 'desired' object is already known, why is it necessary to 'find' it?"[396]

This homunculus is difficult to exorcise. William Randall introduces the metaphor of a "compost heap" for memory, instead of the computer metaphor, for exactly some of the reasons evinced above. Randall is convinced that the computer metaphor cannot bear the strain that our modern understanding of memory has placed on it—and he is right to be so convinced. Then, however, he goes on to posit an inner *gardener*, tending the compost heap, instead of an inner operator tending the computer![397] Still, the compost heap is an improvement: compost heaps are in considerably less need of gardeners than computers are in need of operators. As we have seen above, it is the encoded trace that invites this homuncular explanation. Something that is encoded can be *transformed* by all kinds of processes, but it cannot be *decoded* except by an understanding cognitive agent. Decoding without understanding is not decoding, it is simply re-coding.

---

[395] Dudai 2004, p. 79. Schacter suggests that the medial temporal region may contain "a kind of index that "points to" the locations of different kinds of information that are stored in separate cortical regions" (Schacter 2008, p. 87).

[396] Tulving 1983, p.5.

[397] Randall 2007, p. 622. Perhaps it should be added in Randall's defense that throughout the piece he is supremely aware that his metaphor is a metaphor, and hence is less easily bogged down by conceptual confusions than he might otherwise be.

One way that these problems get started is a notorious equivocation between two different notions of 'information'.[398] 'Information' is sometimes taken to denote *Shannon information*, the maximum quantity of communicated content (based on the likelihood of a particular configuration against the pre-arranged possibilities) that something could have *if* it were interpreted as a message. On the other hand, 'information' is also sometimes used to denote meaningful, semantic, *interpreted* information. "Information processing" explanations of cognition have started to wane in frequency and popularity in cognitive science in recent years.[399] In addition, those explanations which are still offered in these terms have been increasingly clear about what is meant by 'information', and hence less inclined to conflate semantic and non-semantic varieties. Nonetheless, this conflation is still very much in evidence. Memory scientists are at least as prone to this equivocation as any other researchers in cognitive science.[400]

One way to be sure not to conflate these is to be sure to keep the subpersonal elements of an explanation non-semantic. This strategy, however, reveals problems that inhere in the supposed cognitive level of explanation. This mental software level makes sense in a computer, which is only transforming non-interpreted, non-semantic information until it is interpreted by a cognitive user, but it is not clear how such a level could proceed in human cognition. Andy Clark writes,

> The strategy of focusing attention on a kind of disembodied "pure software" level, floating high above the messy material realm, works well when we confront computational systems of our own design. But it works because of the simplifications, regimentations, and neat

---

[398] Malcolm 1977, p. 217; Dreyfus 1972, p. 77. That meaningful content as interpreted naturally by humans in non-prearranged settings cannot readily be brought under the auspices of Shannon information is nicely illustrated by Hutchins (1995, Ch. 5). See also Hutto and Myin 2013, p. 67.

[399] Although witness Gallistel and King: "Most cognitive scientists think about the brain and behavior within an information-processing framework" (Gallistel and King 2011, p. 1). Notably, Gallistel and King's attempted solution of the information equivocation issue involves making the brain an *interpreter* of information (*Ibid*., ch. 1).

[400] Alan Baddeley writes, "Complex thought requires the manipulation of information, and such manipulation demands temporary storage" (2007, p. 151). Cf. Treves 2007, p. 55.

> decompositions *we* artificially impose on the electronic circuitry so as to make it tractable to the process of conscious design in the first place.[401]

This inner cognitive level at which the processes of remembering are supposed to take place is defined not by what we have found at this level, but by what we have not found anywhere else. That is, by the assumptions memory researchers adopt, certain elements and processes of memory—such as traces, retrieval, encoding—must exist. These have not been located at the neurobiological level, and they have not been located at the behavioral level (or in the phenomenal domain, however this fits with the aforementioned levels). And so, if there are such inner processes and states necessary, they must exist at this cognitive level. This, however, is a statement about the necessity of *both* of these things: the collection of processes and states *and* the level on which they will be found. That is, by setting up the investigation of memory in this fashion, researchers have co-posited the need for inner processes and the need for certain inner processes. These mutually depend on one another, and so each in isolation is taken to be essential. In actuality, the inner remembering level is just a way to collect all of the problematic consequences of our assumptions in one basket.[402] Unconscious processes are notorious problem-solvers: they are too good at accounting for whatever needs accounting for.[403] Worse, once we have conceived of remembering in a way that compels us to explain it in terms of unconscious retrieval, the way out of our philosophical difficulty becomes obscure.

---

[401] Clark 2001, p. 100.

[402] David Stern writes, "But the problem of connecting the causally effective physical level of explanation with the cognitively significant intentional level is a restatement of the problem of connecting mind and matter, the central problem of Cartesian dualism, rephrased in the idiom of 'levels of explanation'" (1991, p. 205).

[403] This worry has come up in many contexts, but we might go back to James' *Principles of Psychology*: (p. 163) "The distinction…between the unconscious and conscious being of the mental state…is the sovereign means for believing what one likes in psychology, and of turning what might become a science into a tumbling-ground for whimsies." See also Ayer's remarks against unconscious recollecting as quoted above on page 161 (Ayer 1956, p. 136).

In *The Blue Book*, Wittgenstein addresses this.

> Thus, by the expression “unconscious toothache” you may either be misled into thinking that a stupendous discovery has been made, a discovery which in a sense altogether bewilders our understanding; or else you may be extremely puzzled by the expression (the puzzlement of philosophy) and perhaps ask such a question as “How is unconscious toothache possible?” You may then be tempted to deny the possibility of unconscious toothache; but the scientist will tell you that it is a proved fact that there is such a thing, and he will say it like a man who is destroying common prejudice. He will say: “Surely it’s quite simple; there are other things which you don’t know of, and there can also be toothache which you don’t know of. It is just a new discovery”. You won’t be satisfied, but you won’t know what to answer. This situation constantly arises between the scientists and the philosopher.[404]

It can seem as though there is nothing mysterious about unconscious cognitive mechanisms of memory, because we are familiar with so many other cases of processes that we have slowly come to understand. Many of these are literally inner processes—understanding physiological processes takes time and investigation.

We have already seen that the component processes of memory are sometimes framed in ambiguous terms. The experienced imagery that we sometimes remember with is at times taken to *be* the inner process of remembering, thus conflating inner cognitive processes with inner, experienced, mental processes. Perhaps worse, the “inner” brain processes are sometimes ambiguously referred to haphazardly alongside inner cognitive processes. If one memory scientist describes what it is that is stored and retrieved as “neural firing patterns”, [405] and another describes what it is that is stored and retrieved as “learned information”,[406] and yet another describes what it is that is stored and retrieved as “experience”,[407] the level at which these processes exist is just not being clearly and unanimously articulated. This avoidance of any meaningful convergence on

[404] *The Blue and Brown Books*, p. 23.

[405] Eichenbaum 2007, p. 193.

[406] Roediger et al. 2007, p. 191.

[407] Gabrieli 1998, p. 88.

the supra-neural, infra-behavioral level is indicative of this level's inability to carry out the promises that have been placed on it.

There is another alternative for the "inner" elements that are posited to explain human memory. Sometimes these elements are treated as theoretical constructs, as elements in a model in the same way that there are elements in economic models or ecological models. This is quite different from an unconscious process, and theoretical constructs are certainly neither neural nor behavioral (nor phenomenal). In ways that will be further discussed in the following chapters, this model-element explanation of the subjects of study in the cognitive science of memory is one of the most hopeful. We can certainly model any number of phenomena in this fashion, and these explanatory models really do help us to understand and predict the behavior of complex systems. If held to consistently, this model-model of memory science can save some of the successful and satisfying elements of psychologists' theories while avoiding some of the usual pitfalls. However, the subprocesses of memory as they are standardly characterized do not consistently fit this schema. Theoretical constructs are not *implemented* by the physical, nor are they mechanisms by which ground-level phenomena are *caused*. The axis of the earth's rotation is such a construct, a simple and useful element of a successful explanatory model. To ask what physical stuff or processes implement or realize the earth's axis is simply to be confused about the type of thing it is. Only a young child or otherwise extremely naïve investigator would claim that, or ask whether, the axis *causes* the rotating we are presently doing. Of course the axis does not *cause* the rotation. Rather, the best way for us to simply and successfully understand the rotation of the earth is to posit this axis.[408] Whatever these inner processes of remembering are supposed to be, there is no

[408] Daniel Dennett makes very similar points about theoretical constructs and psychological ascriptions, utilizing the notion of "center of gravity" as an example. This discussion first appeared in Dennett 1981, but can also be found in more detail in Dennett 1991.

place for them. If an element of remembering is neural, behavioral, phenomenal, or an element in an explanatory model, so be it. Each of these has a place in the study of memory. The nebulous and shifting processes of inner, non-conscious, non-model processes and states do not belong.

4.3.5 *Processes with and without neural correlates*

The assumption, nearly ubiquitous in the study of memory, that remembering is the product of hidden processes is composed of two parts. The first is that these cases of remembering are subsumed under process-types at the hidden, cognitive level. This has been discussed in the last section. These cognitive process-types are also taken to be implemented by neural process-types. This claim, that acts of remembering are implemented by particular neural process-types, is the last computationalist assumption that merits investigation and destabilization.

Once more, it can seem like the neural correlate of remembering is the only game in town. It is uncontroversial that the human brain is responsible for human cognitive activity. It is uncontroversial that brains, particular brain regions, and even particular neural mechanisms are necessary conditions for all kinds of cognition. Neuroscientists can inhibit learning in laboratory animals in very specific ways by means of lesions to brain regions or particular drug cocktails, as we have seen above. To doubt whether there is a neural correlate of remembering can at first seem as if it is equivalent to doubting all of science.

However, the assumption that there are neural correlates of remembering is the assumption that there are neural process-*types* that correlate to remembering process-*types*. There is no reason to doubt that there are token-token correlations between particular instances of remembering and particular neural and physiological processes. Indeed, we could only deny this by undermining all of science. For any given instance of remembering, something has happened in the brain and body. This is just another way of saying that we would not predicate ‘remembering’ of an inanimate object or a corpse. None of this guarantees that there are neural correlates of memory in general. That claim is not just a claim about cognition and its relationship to the brain, it is a claim about what kind of thing remembering is. There are many cognitive activities humans engage in for which we would not expect to find neural or physiological correlates. There are some cognitive activities that we do, at least to some extent, expect to find such correlates. This is a familiar pattern, but it bears repeating: to assume that memory belongs to the latter and not the former without argument is to invite disaster. It is not surprising that we make this assumption. In The *Brown Book*, Wittgenstein elaborates on the human tendency to see psychological statements about what is possible as indicators of an underlying state.

> The same tendency shows itself in our calling the ability to solve a mathematical problem, the ability to enjoy a piece of music, etc., certain states of the mind; we don't mean by this expression 'conscious mental phenomena'. Rather, a state of the mind in this sense is the state of a hypothetical mechanism, a mind model meant to explain the conscious mental phenomena. (Such things as unconscious or subconscious mental states are features of the mind *model*.) In this way also we can hardly help conceiving of memory as of a kind of storehouse. Note also how sure people are that to the ability to add or to multiply or to say a poem by heart, etc., there *must* correspond a peculiar state of the person's brain, although on the other hand they know next to nothing about such psycho-physiological correspondences. We regard these phenomena as manifestations of this mechanism, and their possibility is the particular construction of the mechanism itself.[409]

But this indication of an underlying state is belied by particular cases of those cognitive phenomena which obviously are not suited to such. Take phenomena such as expectation or hope. The ability to hope is clearly cognitive—non-cognitive entities cannot hope, and any hoping I manage to do will be with or by means of my cognitive capacities. However, any search for the "neural correlates of hope" will be doomed. Although each individual act of hoping surely brings along some neural processes, these individual acts are constituted by such a wildly rich variety of processes and actions that the folly in seeking a general neural process-type is immediately

---

409 *Blue and Brown Books*, p. 118, original emphasis.

apparent. One can hope by speaking. One can hope by inner monologue. One can hope by merely shrugging one's shoulders in the face of expressed cynicism. It would be no less remarkable to find an "underlying" neural correlate of these cases than it would be to find the neural correlate of checkmate, agreement, or art appreciation.[410]

This is already sufficient reason to call the automatic status of the neural correlate assumption into question. If there are some cognitive capacities that are the wrong sort of thing to go looking in the brain for their correlates, then what is needed is a way to establish whether memory is one of these. There are also mental states that do have correlates in the nervous system and physiology of the subject—we might take hunger, for example. The experience of hunger, at least, is mental in nature, and there are very specific inner physiological correlates of this sensation. These physiological processes were not always known to us: we gradually came to understand them better and better. The neural processes of remembering are treated as though they fit this pattern.

However, there is at least one compelling reason to classify memory with hope rather than with hunger. It is easy to identify one feature common to all sensations of hunger: the experienced feel of the sensation of hunger. As we have already seen, philosophers and memory scientists agree that there is no such phenomenological common feature among cases of remembering. Instead "we find that what connects the cases…is a cast number of overlapping similarities, and as soon as we see this, we feel no longer compelled to say that there must be some one feature common to them all."[411] The psychologist Jens Brockmeier argues that memory scientists have,

---

[410] This does not always keep neuroscientists from trying. Occasionally, the overly ambitious researcher in neuroscience goes looking for the neural correlate of trust (Babiloni and Astolfi 2012) or the neural correlates of pride and gratitude (Zahn et al. 2009).

[411] Wittgenstein, *Blue and Brown Books*, p. 87.

generally speaking, severely exaggerated the unity demonstrated by the phenomena of memory. As Brockmeier points out, there are some languages that do not even have a single word or concept that isolates what we call memory from other cognitive abilities such as thinking and understanding.[412]

Memory also obviously ranges over nearly the whole of human activity. If there were neural correlates of memory, these must in almost every case be co-extensive with the neural correlates of a great many other cognitive capacities. William James made a similar point about language and localization in the brain.

> It is plain that the faculty of spoken language alone is so complicated as to call into play almost all the elementary powers which the mind possesses, memory imagination, association, judgment, and volition. A portion of the brain competent to be the adequate seat of such a faculty would needs be an entire brain in miniature,—just as the faculty itself is really a specification of the entire man, a sort of homunculus.[413]

While memory science has, in the century since James wrote, determined tight relationships between certain brain regions and certain speech capacities, this has not been the discovery of "the seat" of language in the brain. For example, it was already accepted when James was writing that lesions in Broca's area or Wernicke's area often (but not always) result in particular aphasias and other speech and language deficits. This has been confirmed and elaborated in the last century, but no neural mechanisms of language have been localized. These brain regions do not even constitute necessary conditions for speech or verbal understanding.[414] Among cognitive

[412] Brockmeier 2010, p. 6.

[413] James 1890/2007, p. 29.

[414] The complete removal of Broca's area sometimes results in no speech pathologies at all (Plaza et al. 2009).

neuroscientists, the question of whether language abilities can be localized in the brain remains open.[415]

The capacity to remember is no less complicated or wide-ranging than the capacity for language. Remembering regularly involves imagination, judgment, association, action, and language. Hence, although cognitive scientists sometimes speak as though memory's localization in the brain is right around the corner—and notably, this claim is more common among those who study memory but not the brain, or the brain but not memory, than it is among those who study both—it is difficult to know what such localization could even look like. These considerations have become more widespread among cognitive scientists in recent years. Katharina Henke argues for abandoning neural bases as distinguishing features of memory systems.[416] John Gabrieli doubts whether the neural correlates of language and the neural correlates of memory can be separated.[417] Of semantic memory in particular, Sharon Thompson-Schill writes that "the search for the neuroanatomical locus of semantic memory has simultaneously led us nowhere and everywhere.[418] Matthew Brett and his colleagues have argued that functional localization in the brain cannot advance at all without significant conceptual revision to the very idea of localization and function.[419]

---

415 A century after Broca, Wernicke, and James, several cognitive neuroscientists still doubt this (for example, see Wilmes and Poeck 1993), and even those who today are optimistic about localization are cautious and modest in their claims (see Plaza et al. 2009).

416 Henke 2010.

417 Gabrieli 1998.

418 Thompson-Schill 2003, p. 288.

419 Brett et al. 2002.

There are some cognitive processes and phenomena that are not amenable to localization (and there are some that are so amenable). Given the absence of unifying features, the wide-ranging and complex interrelations between memory and other capacities, and the stubborn difficulties neuroscientists have encountered in trying to establish neural correlates of memory,[420] it must be worth revisiting our assumption that remembering has neural correlates. Remembering bears all the marks of a cognitive capacity that is not implemented by a particular neural mechanism. It should be remembered that this is not to deny token-token neural correlates: every particular act of remembering is also correlated to a particular collection of neural and physiological processes (as is every token act of stalemating, correcting errors, hoping, agreeing, or understanding), but there is good reason to hold, at least until proven otherwise, that no type-type correlations between neural processes and remembering processes will be discovered.

---

[420] Remember the passage from David Sweatt, quoted in section 2.3.3 above, in which Sweatt deduces just how the neural mechanisms of retrieval must be in order to accord with the assumptions in the cognitive science of memory, and then expresses some dismay at just how "*quite mysterious*" these seem to be. From a biological perspective, the resulting mechanisms seem nearly implausible (Sweatt 2007, p. 212). Other researchers concur—for example, admitting that the way memory is conceived to operate "*must mean* that some trace of the memory exists in the brain but it has been difficult to find" (Thompson 2007, p. 200, emphasis added).

CHAPTER V

ALTERNATIVES TO STORAGE

> [Memory researchers] know that memory is far more dynamic than our models have typically allowed. If we are to make further progress in understanding what goes on when we mentally re-visit the past, we must begin to build these dynamics into our models. I would argue that the surest way to do this is to step outside the reigning paradigm. This may strike some as a frightening prospect—abandoning the known in favor of the unknown is always a bit scary. However, we should not let fear of the unknown bind us to a way of thinking that increasingly appears to have reached a dead end. The way forward involves facing up to the complexity of memory and the processes involving the acquisition and constant transformation of knowledge as a function of experience.
>
> Lynn Nadel[421]

> "The faculty [of memory] does not exist absolutely, but works under conditions; and the quest of the conditions becomes the psychologist's most interesting task."
>
> William James[422]

The foregoing chapters have been largely critical. This is needed in memory science—the computationalist assumptions that were outlined in chapter two and shown to be problematic in chapters three and four deserve careful criticism. As a result of this criticism, much of the structure of mainstream models of memory has been undermined—not just in the present work, but in a wide variety of work in philosophy and cognitive science that has been synthesized above. Because these models have enjoyed such widespread acceptance,[423] one legitimate concern is whether these simply constitute the only game in town. If remembering is not a hidden cognitive process, implemented by a neural "remembering" process, in which an encoded memory trace is

421 *Consolidation: The Demise of the Fixed Trace*, p. 181.

422 *Principles of Psychology*, vol. 1, p. 3.

423 Furthermore, it was my aim, especially in section 3.2 above, to show that even revisions to the standard models have in many ways still accepted the problematic underlying assumptions.

stored and retrieved, what could it be? Memory science takes something like this description to be the object of what has been studied for the last several decades—does undermining the standard model mean undermining all of memory science? It does not. Much work in the cognitive neuroscience and cognitive psychology of memory has constituted real advances in the field, and any model of memory must not deny this. Modeling memory in a way that does not include the computationalist assumptions above is a difficult task. In the following chapter I will illustrate a few ways in which the study of memory can proceed, and in some cases is already proceeding, without these assumptions. In so doing, I will attempt to steer between, on the one hand, the dismantling of memory science as we know it and, on the other hand, theoretical revisions that are ultimately only cosmetic.

Psychologist Jens Brockmeier surveys many of the above issues—especially the empirical issues, but also some of the philosophical issues—and he comes to a surprising conclusion. Rather than attempting to tweak existing models one more time to account for difficult data, or to get around an impasse by transmuting one concept into another, Brockmeier simply sees memory, as a meaningful and even remotely unitary object of study, as irredeemable.

> I have put forward the argument that the concept of memory, even in its traditional heartland of cognitive psychology and neuroscience, is about to lose its long taken-for-granted ontological gravity, that its semblance of a rectified entity which, like the lung or the heart, has a safe and sound place in the world or, at least, in the human head, is dissolving. With its unparalleled rise over the last decades, the neuroscience of memory is about to lose its original subject.[424]

Brockmeier's arguments that the "memory crisis" runs deep, and extends across many fields, are impossible to dismiss. He highlights many of the same difficulties that we have already encountered in chapters three and four above. Furthermore, Brockmeier is certainly not claiming

---

[424] Brockmeier 2010, p. 25.

that the individual phenomena that researchers presently study are no longer to be studied. Rather, he surveys these difficulties in the study of memory and concludes that the notion of memory is "in the midst of dissolving," that we are simply beginning to realize that there is no subject called "memory" to study.

As far as reactions to the memory crisis go, Brockmeier's might be contrasted with the reaction of Henry Roediger and his collaborators, which was introduced above.[425] Roediger is witness to many of the same conflicts and confusions that Brockmeier draws on, but he and his collaborators instead see this as a good reason to clarify the concepts researchers employ. Roediger expresses optimism about the future of the study of memory, but only if researchers can pin down these concepts with sufficient clarity and unanimity. He goes on attempt just this, and offers the concepts of encoding, persistence, and retrieval for clarification and consensus. As we have already seen, this approach is representative of many in the sciences of memory, in which a few terms have been revised, but in which the same basic, underlying, computationalist framework persists. It was demonstrated above that much of the work done by Roediger's cohort is in important ways carrying on as usual, despite their best intentions to do otherwise.

Brockmeier surveys the problems of memory science and foresees decoherence, claiming that the conceptual framework within which memory has traditionally been studied has been undermined so profoundly that no study of memory, as we know it, can survive. [426] Brockmeier is not alone, nor is he the first to conclude that memory is not unitary enough to be a cohesive subject of research and analysis. Martin Deutscher—the same philosopher who co-authored the

---

425 The following especially pertains to the contributors to the volume *Science of Memory: Concepts* (Roediger et al. 2007). These contributors are many, and they comprise a significant portion of the influential researchers working in the cognitive psychology or cognitive neuroscience of memory in the 21st century.

426 It is worth noting that Brockmeier also draws heavily on literary, artistic, technology-based and other recent cultural treatments of memory from a wide array of sources (Brockmeier 2010).

influential account of memory above—decades after his attempt to give an account of memory instead wrote that there is no such phenomenon to account for.[427] But somewhat paradoxically, these same data have pushed different researchers in opposed directions. Roediger and his collaborators, surveying many of the same difficulties Brockmeier or Deutscher note, urge clarification and consensus, hoping that this traditional framework—perhaps modified in its details—can serve to unite and aid a study of memory. It should be clear, by now, that there is something in each of these perspectives that marks them as kindred to the present project. Nonetheless, the positive view of successful accounts of memory that I want to offer is just one that forges a path between the "giving up" prophesied by Brockmeier and the "carrying on" embarked upon by Dudai and his cohort. If the study of memory is to survive as a multidisciplinary whole—and I believe that it can, and will, at least insofar as any multidisciplinary study can be treated as a whole—then we do need to investigate the conceptual framework implicit in our approaches to the subject. However, this will just involve the whole-hearted rejection of the computationalist assumptions that have been generating such difficulty. In this chapter, I will sketch accounts of memory that reject these underlying assumptions but which do not lead memory, as a phenomenon worthy of study, to dissolve altogether.

5.1 Resources from anti-computational developments in cognitive science

There are several movements in cognitive science that offer particular helpful resources for this new conception of what it is to remember. The first is a tradition of research and analysis surrounding the development of artificial neural networks. Another is the multi-disciplinary and variegated collection of research programs that go under the heading *dynamic systems*. Yet another

[427] Deutscher 1989. Malcolm also sometimes sounds as though he in inclined to conclude the same in *Memory and Mind* (1977).

is the even more various family of research programs that are associated with situatedness, embeddedness, and embodiment. Many memory researchers have at least casually drawn on connectionism and the connectionist tradition for ideas about restructuring our models of memory,[428] and some researchers have also looked to the tools of dynamic systems research for new and better ways to characterize remembering and forgetting.[429] The burgeoning studies of embodied, embedded, or situated cognition perhaps tend to have the most overlap with memory science and research:[430] as evinced in the repertory example above, there are at least a few minor research programs that explicitly combine these research traditions with the study of memory. Nonetheless, each of these has important and ongoing lessons to offer to the memory researcher who seeks to avoid the criticisms above.

The first application of these to problems in memory science has already been noted, in the inconsistent triad from section 3.2.1. Most cognitive scientists of memory do now characterize the trace, or engram, or stored information, in terms of a distributed item in a network.[431] Most cognitive scientists of memory also still characterize the trace as both isomorphic and representational (when they characterize the trace at all). However, researchers who study neural networks are, by and large, very uncertain as to whether there are any connectionist vehicles that are appropriately described as representational or even isomorphic to the stimulus reproduced by

---

[428] Sutton 1998; McClelland 2000; Norman et al. 2007.

[429] Kryukov 2008; Schöner 2009.

[430] By lumping embodiment, embeddedness and situated cognition together, I do not mean to commit to any particular stance concerning their unity. I agree with Kiverstein and Clark (2009) that the jury is still out regarding distinctions between these, and that these distinction questions are important. However, it is presently difficult to draw the boundaries with any precision, and I will here treat these views together.

[431] Ramsey mentions the cognitive science of memory in particular as a locus of the notion of "tacit representation", because of the connectionism's "enormous influence on the cognitive and computational neurosciences" (Ramsey 2007, p. 159).

the network. At best the reproduction of the network may in some cases be representational of and isomorphic to certain input, but even in these cases there are no features internal to the network, weights, or activation states that map to the features of the input or output.

Neural network research has demonstrated that there can be systems that identify, transform, and replicate very complex patterns without storing features of these patterns in any straightforward fashion. In 1991, the mere demonstration of such systems was an exciting and interesting development.[432] Two and a half decades later, such systems have become unremarkable for researchers who work on connectionist systems. Debates concerning the representational nature, or otherwise, of the states of the network have quieted in the last decade.[433] In part this is because connectionist researchers have simply gone on not invoking representation, and cognitive scientists working in other areas, and employing representations, have casually adopted connectionist language. Memory researchers are, in many cases, among the latter. Any memory researcher who takes the "neural change" accompanying a past event to be a connectionist-style change in a neural network should also rethink any description of this resultant change as "representational" or "isomorphic". Neural network researchers manage to give very rich, detailed, rigorous descriptions of the inputs, outputs, activities, and internal states of their networks. If memory researchers seek to rid their models of a "fixed trace"—and as we have seen many times above, several memory researchers explicitly avow that they seek this—they should study the characterizations on offer from connectionism.

---

432 Jeffrey Elman emphasizes that in a network which reproduces words, "there are no representations of words in isolation" (Elman 1991, p. 378). Whatever internal states result from a given input are always a summation of the whole prior state of the network, plus the reaction to the stimulus.

433 A simple search reveals about twice as many articles on "connectionist representation" between 1994 and 2004 as there have been between 2004 and 2014. The peak seems to have been around 1997.

Another helpful lesson from neural network research concerns the separation of algorithm and data. In part due to the computer metaphor, it can be difficult to conceive of remembering as anything but a series of processes (retrieval, consolidation, etc.) that are acting *on* something (the trace, the engram, the memory, etc.). If we have reason to reject inner, modular processes like retrieval, and reason to reject traces, it may seem as though there cannot be any space left for interestingly complex processes at all. Neural networks illustrate, however, that this tendency to equate complex information-preserving processes with processes of algorithms and data is just another vestige of computationalism. Artificial neural networks simply do not exhibit the kind of separable components found in traditionally computationalist models.[434] In highlighting connectionism's lessons for cognitive science, Andy Clark draws out attention to this demonstration.

> One feature of [connectionist] models was the apparent collapse of the data/algorithm distinction itself. The connection weights, in such models, act as both knowledge store and knowledge-manipulation algorithm. If real neural computation is indeed anything like connectionist computation, the standard notion of an algorithm as a recipe for acting on an independent data set also seems strictly inapplicable.[435]

Once more, researchers studying neural networks give detailed and rich descriptions of the states and processes of the network without recourse to any *items* or *data* that are the objects of algorithmic processes. We have already seen that memory scientists are presently engaged in a debate about the *reconsolidation* of memories. One of the factors that has made this debate contentious and difficult to clarify is that there is increasing evidence for *continuous* reconsolidation. That is, there was once posited to simply be a phase of consolidation following

---

[434] "Although neural networks have architectures that can be depicted as separate systems, the are—at their core—complex, reentrant, densely interconnected, complex systems that violate the separability assumptions of information processing" (Spencer et al. 2009, p. 89).

[435] Clark 2001, p. 97.

the encoding phase (but perhaps before the "storage" phase). Then it was discovered that once-stable memoires can under certain conditions become labile again, and a "reconsolidation" phase was posited.[436] Finally, it seems that memories are actually more or less continuously labile, or at least that they undergo changes as a result of almost any associated activity. Memory scientists are having a difficult time making sense of a "phase" of remembering that is apparently never-ending.[437] This confusion, however, is in part due to the continued attempt to separate the process from the product. What is slowly becoming clearer to those working on reconsolidation is that there is no longer conceptual space for an isolable *object* of consolidation if reconsolidation is continuous. But for connectionism's collapse of data and algorithm, this could be seen as an insurmountable difficulty. If memory scientists seek a way to characterize this continuous reconsolidation, they will do well to consider the ways that connectionist networks lack any clear separation between data and process.

Andy Clark also advocates replacing our traditional concepts of mind with concepts from dynamical systems theory, because in his words the image of cognition as computation is a "throwback to the idea of the brain as, in essence, the seat of a fundamentally disembodied kind of intelligence."[438] Dynamical systems research is unified more by its methods than by its subjects. Strictly speaking, "dynamic systems"—systems that evolve over time—include almost every system of interest. The dynamical systems research program, however, consists in those projects which emphasize certain types of interdependence among the variables in the system, as well as certain types of mathematical and modeling tools to characterize these developments and

---

[436] See McGaugh 2000 for a survey of these developments.

[437] Nadel 2007.

[438] Clark 1998, p. 98.

systems. Whether the study of neural networks is merely a special case of the study of dynamic systems remains an open question.[439]

One resource from dynamical systems research that deserves further attention from memory scientists is the large and powerful array of modeling techniques that have been developed. Since the theory is based in method, models are, in a sense, the very center of dynamical systems theory. It was demonstrated above that the theoretical constructs of explanatory models are one of the candidates for elements of memory models that do not run afoul of the philosophical and empirical objections from chapters three and four. If examples of these theoretical constructs are limited to centers of gravity and axes of rotation, it might seem as though such elements are not robust or powerful enough to do the explanatory work that cognitive scientists seek in explaining human memory. The models of dynamical systems theory, though, are impressively robust and powerful, and they in many cases describe very complex behaviors of very complex systems with great precision and rigor. Furthermore, dynamical systems explanations of cognitive phenomena (e.g., the collected articles in Spencer, Thomas, and McClelland 2009) present a challenge to a popular view of psychological explanation. Robert Cummins noted that psychology tends to designate "effects" rather than laws, and that these effects do not explain the phenomena they subsume, but rather stand in need of explanation themselves.[440] Dynamical systems "explanations" may not satisfy Cummins' demand, for they merely consist in exhaustive descriptions (and predictions) of the system's actual and counterfactual behavior. Nonetheless, at least some psychological phenomena seem to be explained by dynamical models, and it is not clear in these cases what sort of explanation could still be lacking. There already

[439] Increasing numbers of researchers seem to agree that, at the very least, these two domains overlap significantly with one another, and advocate further integration (E.g., McClelland and Vallabha 2007, Spencer et al. 2009).

[440] Cummins 2000.

exists work that gives accounts of remembering phenomena in terms of dynamical models, but there remain many opportunities for further work along these lines.[441]

Lastly, there are several resources and strategies that have recently become available from work in embodiment and embedded or situated cognition. Two recurring themes are especially important. First, there has been a collection of recent work, especially from distributed, situated and embedded cognition traditions, that emphasizes how little we actually need in-the-head representations and stored knowledge in order to accomplish most of what we accomplish in our cognitive lives. Hutchins' *Cognition in the Wild* is an early and representative example of this: Hutchins explicitly allows for representations stored in individuals' memories, though he does not study or characterize these. Rather, he shows that these internal, stored representations play a remarkably small role in cognition as it happens in the real world. Hutchins' case study is the navigation of a navy frigate—a task that does not at first glance seem unsuited to mental representation and computation. Nonetheless, Hutchins makes a convincing case that even these highly intellectual and representational problems are simply not solved by means of traditional, in-the-head, computation-style cognition. The implications for memory are very different than the implications from a theory like Stephen Kosslyn's account of perception, according to which we need stored visual representations even to recognize the objects in the world around us.[442] If Kosslyn's more traditional account is right, memory must produce something that is at least functionally similar to stored items on a relatively constant basis. If Hutchins is right, there are relatively few occasions in our cognitive lives where we rely on anything like internal storage. This shift changes the problem space for the memory researcher. Rather than assuming that we

---

[441] Spencer et al. 2009; Kryukov 2008.

[442] See footnote 389 on page 188 above. Burge sometimes characterizes perceptual memory and recognition capacities in similar ways (see, e.g., Burge 2010, pp. 446-448).

must constantly match the world to our stored memories even in basic problem-solving, the distributed and situated traditions only ask that very particular memory tasks can be performed when needed. Much of our cognitive work can be accomplished by and in our environment and interactions.

Second, there has also been a collection of recent work, especially from radical embodiment and enactivism, that emphasizes how much can be accomplished in the absence of stored representations. This theme in cognitive science runs back at least as far as the subsumption architectures of Rodney Brooks[443] (1993) and the embodied mind of Francisco Varela and colleagues (1991),[444] but very recent treatments like that of Chemero's *Radical Embodied Cognitive Science* (2011) or Hutto and Myin's *Radicalizing Enactivism* (2013) continue to demonstrate that many complex cognitive acts can be successfully performed without storing representations at all. The basic lesson from Van Gelder's example of the Watt governor, that sometimes problems that look like they would require stored and retrieved representations can in fact be solved in elegant, nonrepresentational fashions,[445] is one that memory scientists would do well to take heed of before they too quickly assume that the problems solved by memory must be solved using representations. Even the memory models, like those from Dudai and colleagues above, that attempt to account for non-computational aspects of human remembering often fall back on algorithmic approaches in spite of themselves. The lesson from these embodied approaches to mind and cognition that algorithms are not the only game in town, and that a carefully constructed interdependent and dynamic model can accomplish cognitive tasks better, in

---

[443] Brooks 1993.

[444] Varela et al. 1991.

[445] Van Gelder 1995.

some cases, than a computational separation of data and algorithm leaves memory researchers with more options than they might think.

5.2 Features in the absence of modularity

It has been argued that, in many ways, the human capacity for remembering does not lend itself to the distinctions that traditional memory science has imposed on it. The distinction between remembering process and remembered item does not bear the weight of empirical results. The distinction between memory systems cannot cleanly divide remembering behaviors nor neural mechanisms. The distinctions between subprocesses such as encoding and retrieval are useless in describing the phenomena. The very notion of a distinct, inner process of remembering—either cognitive or neural—has been undermined along with the separation of remembering from other cognitive capacities. Almost every module in the traditional account of what it is to remember has been shown to be considerably less modular that was assumed.

Taking all of these together it might seem as though remembering is now condemned to be a sort of black box, theoretically speaking. It is a process without parts, and these parts may overlap with nearly any other psychological process. Remembering produces certain behaviors, but these do not belong to natural divisions or systems. If the inner workings of memory can no longer be intelligibly described, what could be left to say about remembering? The resultant picture of memory can seem implausibly holistic and resistant to explanation. In what follows, I aim to re-establish room for features in remembering, even in the absence of the modularity that computationalism engenders. An important part of the Wittgensteinian lesson from section 4.1.1 above was that remembering really does happen in many ways. Even if these do not divide into distinct systems, subprocesses, or traces and hidden processes, memory researchers can certainly model, predict, describe, and explain features of these ways of remembering.

5.2.1 *Generative remembering*

One of the clearest and well-reasoned accounts that attempts to resolve some of the fundamental puzzles and difficulties in modeling memory can be found in philosopher Kourken Michaelian's recent article on what he calls generative memory.[446] Although Michaelian borrows much from the traditional framework for understanding memory, he is sensitive to some of the ways that these assumptions are contravened by the actual workings of memory. Michaelian's account is useful as what might be the best instance of the stratagem propounded by Dudai and his collaborators above: the generative memory model keeps the basic framework of a stored trace that is encoded and retrieved, but clarifies and revises these concepts in an empirically-informed fashion. One reason this account is worth studying, then, is simply the fact that it exemplifies a best-case scenario for an approach that I ultimately deem insufficient. Another reason to start with Michaelian's account, however, is that I aim to show that the model that results from Michaelian's refinement is actually no longer dealing in the storage and retrieval of traces.

Michaelian notes that there is an increasing and convincing body of evidence for the fundamentally *constructive* nature of remembering found in the science of memory. The extent to which normal rememberers *construct* rather than *preserve* the past is staggering.[447] Confabulation—the "false" memory of events that did not actually happen—is easy to induce in healthy, adult subjects.[448] Schematic but inaccurate representations of past stimuli are recognized

[446] Michaelian 2011a.

[447] Much of this was surveyed above—the ways that cues and contexts can interfere, positively and negatively, with our ability to remember are myriad. See the introduction to chapter three for a brief treatment.

[448] Zaragoza and Lane 1994.

more readily than veridical ones.[449] Past scenes are systematically remembered as including a broader frame than they in fact did.[450] The most commonplace instances of remembering the past seem to involve profound abstractions, reconstructions, and distortions.

Michaelian points out that this pervasive and ubiquitous construction is accepted by memory scientists, but ignored by philosophers. He suggests that this is perhaps because philosophers are inclined to see such phenomena as mere contingencies of the biological implementation of memory.

> Some might argue that the fact that encoding in human memory is constructive is of little significance to the philosophical theory of memory: construction at encoding, they might suggest, merely reflects contingent features of the way in which memory is implemented in our species; the theory of memory thus need not take it into account, but should, instead, focus on what there is in common between constructive human memory and other, non-constructive possible types of memory.[451]

Michaelian goes on to argue that these are not contingent features of human memory. For each of the ways that memory is constructive, he argues that this "is not the exception to the rule; it is the rule."[452] This is right. Paying attention to the way that human remembering actually works makes it implausible to continue to characterize these constructions as exceptions. Reconstructing, distorting, distilling, and adding to the past is what remembering does.

In response, Michaelian's theory of "generative remembering" incorporates these into what it is to remember. He suggests that remembering just is when the reproduced content "does not

---

[449] E.g., the "the caricature effect" of Tversky and Baratz (1985) demonstrates that recognition succeeds more readily with a caricature than a photograph. There is a tradition of similar experiments with "gist memory" (starting with Fillenbaum 1966).

[450] Intraub et al. 1992.

[451] *Op. Cit*., p. 326.

[452] *Ibid.,* p. 333.

go too far beyond" the past event or experience it is reproducing.[453] Like many of the revision-minded memory researchers we have encountered above, he puts this in terms of a distributed memory trace that is distorted at encoding and distorted again at retrieval. These distortions, however, are just part of what makes up memory.

> Thus we need not and should not say that the subject does not genuinely remember the scene because she did not see the whole of it; we should instead say that she remembers the scene even though she did not see the whole of it.[454]

Michaelian makes a convincing case that we resist this sort of adoption of the constructive processes of memory because we tend to think of construction as unreliable and mismatched with the past. However, as he points out, construction is typical in remembering, and "does not typically result in mismatches—constructive memory is reliable" (*Ibid.*). That is, the philosopher who takes "remembering" to be a matter of a preservation process that is contingently prone to mistakes errs. We have already seen many empirical results that demonstrate this. Human remembering is not essentially a matter of preservation and replication, which then gets clouded by error. Rather, construction, reconsolidation, and interpretation go all the way down; by all accounts in modern memory science, these are part of what human remembering is.

Michaelian is right to build this construction into the account of memory, even though this has counterintuitive or revisionist consequences. On Michaelian's account, memory is essentially generative: a subject can successfully remember what she did not experience, provided that this generated content does not exceed certain limits. Michaelian admits that the boundaries dividing remembering from seeming to remember are vague, and that this vagueness cannot be eliminated.

> The proposed conditions on the adequacy of a trace are vague as they stand. I suspect that this vagueness is ineliminable: there is a difference between remembering and merely

---

453 *Ibid.*

454 *Ibid.*, p. 334.

> seeming to remember; but there is no reason to expect that we can draw the line with much precision. If a subject "remembers" something that is not the case, we know that she merely seems to remember. If she "remembers" something that is the case, and if little content generation occurs along the relevant causal chain, we know that she genuinely remembers. If she "remembers" something that is the case, and if massive content generation occurs along the chain, we know that she merely seems to remember. We cannot say more than this (*Ibid.*, p. 335).[455]

Michaelian is right that construction is no contingent accompaniment to the essentially preservative core of remembering, and he is right that constructive remembering very often yields successful, everyday performances of remembering. If we make these constructive qualities part of what it is to remember, it must follow that we will be unable to draw a sharp line between remembering and seeming to remember. In the *Blue Book*, Wittgenstein warns us that "the difficulty in philosophy is to say no more than we know."[456] In this regard as well, Michaelian's account of memory is admirable. The indeterminacy that Michaelian asks us to accept as ineliminable is, I take it, an instance of the psychological indeterminacy that, according to Michel ter Hark, Wittgenstein asks us to describe and accept as is.

Michaelian's generative memory heeds many familiar lessons, both from the findings of modern memory science and from Wittgensteinian insights about psychological indeterminacy. However, this account remains within the standard storage-retrieval framework, albeit just. In name, Michaelian's account maintains the memory trace; it also maintains encoding, storage, and retrieval. In practice, Michaelian's theory strains the conceptual limits of these elements. The generative trace is distributed and constructed (and reconstructed), encoding is constructive rather than translational, the stored trace is described as a dispositional memory, and retrieval is constructive. Michaelian characterizes one of the necessary conditions for remembering P thus:

---

455 *Ibid.*, p. 335.

456 *The Blue and Brown Books*, p. 45.

"the causal chain goes continuously via a (distributed) memory trace with the content P (or something sufficiently close to P)."[457] The "sufficiently close" accounts for content generation; a subject can remember scenes she did not see because her memory trace, by this framework might not include some of the things she can rightly be said to remember, and in turn the memory trace can include some things (that she can rightly be said to remember) that did not really happen.

This treatment of the encoding and retrieval of a memory trace, however, cannot sustain the kind of accepted indeterminacy Michaelian is advocating. Michaelian allows for content generation during at least two phases, the encoding and the retrieval of the trace; but establishing the distinction between encoding-generation and retrieval-generation will be no conceptually clearer than establishing the distinction between encoding failure and retrieval failure. Indeed, the former is even less plausible: the encoding-generated content could only be individuated, in principle, as that content which was not experienced and would be retrieved even without any retrieval-generated content. Conversely, retrieval-generated content can only be individuated as that content which was neither experienced nor encoding-generated but is retrieved. To isolate either of these, we would already have to have individuated the other. No test could determine that some content was retrieval-generated before first determining the surrounding encoding-generated content, and no test could determine that some content was encoding-generated before first determining the surrounding retrieval-generated content.

This is not the worst difficulty: although Michaelian characterizes the constructive nature of memory processes in terms of encoding and retrieval, these subprocesses could perhaps be collapsed without difficulty. His formal statement of the necessary (and jointly sufficient conditions) for remembering P make no mention of encoding or retrieval (only "reliable memory

---

[457] *Op. Cit.*, p. 335.

systems" and "appropriate relations"). His account does, though, maintain a crucial role for the memory trace. The trace is partly constructed and partly preservative of the past, but it is a trace no less, according to this account. Michaelian insists that the trace itself has "the content P (or something sufficiently close to P)." That is, the neural trace has features that are isomorphic to the remembered fact or event, even though it also has features that are not so isomorphic.

This is the generative account's undoing as it presently stands. If the trace has some features that are isomorphic to what really happened, and some features that are isomorphic to what is remembered but *not* to what really happened, how can the vagueness of the trace be a matter of strict indeterminacy? That is, as long as the memory trace has structured content and is partly isomorphic to the past event, then there is a basis, at least in principle, for distinguishing between what is "truly" remembered and what is "falsely" remembered, which is just what Michaelian is trying to exclude. Michaelian cannot admit that the content of the trace is partly isomorphic to the past event (or, for that matter that the remembered "representation" is partly isomorphic to the content of the trace) and at the same time insist that there is no principled distinction available, in principle, between that which is genuine remembering and that which is merely seeming to remember. Admitting an encoded trace that has structure (or "content", as Michaelian prefers) [458] that is even partly isomorphic to the past or the future remembering entails the availability of a principled line between matched and non-matched structure (or content), which allows for a principled distinction between remembering and seeming to remember. Remembering cannot be essentially generative and at the same time consist in the encoding and retrieval of an isomorphic trace. If the trace has content, then there is some content, in a non-

---

[458] Michaelian treats the trace (as well as the past experience and the present remembering) as representational, but the problem I am outlining here only requires isomorphic or feature-matched content of any kind.

indeterminate sense, that may or may not match the past or the remembrance. If our model of memory is to include such a trace, then it cannot be essentially generative.

Alternatively, our account of memory can remain essentially generative if we abandon the notion that the memory trace has to have the same or similar *content* as the remembered item. We might take Wittgenstein's proposals seriously. Even in terms of content, the "content" of the remembrance can match the "content" of the past experience even though there is nothing in the interim that bears content that matches either of these.[459] If this is so, then there need not be any distinction available between that remembered content that "really did" match the trace and that remembered content that did not, since there is no longer any trace content to be matched. William Ramsey suggests a similar resolution to the indeterminacy of content mere isomorphism yields, citing the particular circumstances of the cognitive system's use of the isomorphism as the only possible fixer of content.[460] Although Ramsey does retain the language of inner "representations" and isomorphic structures, he admits that such structures cannot alone suffice for fixed content. For reasons given above, there can be no inner neural trace that exhibits Ramsey's s-representation of the past, but the content-fixing properties of the cognitive systems present circumstances can provide isomorphisms with the psychologically indeterminate past in just the way Ramsey is suggesting. With the removal of the threat of the structural standard imposed by the trace,

---

[459] The indeterminacy of match between remembrance and past event can be fundamental in a way that the indeterminacy of match between trace and remembrance cannot, since the trace has an established, determinable structure, which can be individuated as *the* structure of the trace, in a way that the past does not. Gottfried Vosgerau offers an argument that the inactive trace cannot have any content in a meaningful sense because of the fundamental indeterminacy of pure syntax (Vosgerau 2010). This strikes me as an instance or expression of what Ned Block called "The Paradox of the Causal Efficacy of Content" (Block 1990, p. 138), but both arguments focus on representational content in a way that is only orthogonal to similar concerns about mere isomorphism.

[460] Ramsey 2007, §3.3.1.

remembrances can be fully generative in all of the ways that Michaelian has good reason to believe they should be.

This generative memory, in which the remembrance matches the past, to a certain extent but not perfectly and not by means of a matching trace, provides a good basis for models of memory that are truly free of problematic computationalist assumptions. Some phenomena will be remembering and some phenomena will be seeming to remember, even though there is no modular distinction between these. Expressed remembering will match the past event that is remembered in certain ways (and not in others), even though there will not be any rubric for deciding the extent of this matching. The subject may remember what she has not seen, but only if she also remembers what she has seen, and no investigation of her brain will determine the difference in these.

5.2.2 *The trace-system and continuous reconsolidation*

Michaelian's account requires that the "causal chain" of memory goes "continuously via a…memory trace with the content P" and "continuously via a reliable memory system". The trace, however, cannot have content that does or does not match with other stages in the remembering process. Thus, this *trace* requirement and this memory *system* requirement collapse. The remembering *process* no longer has any clear boundaries of demarcation from the remembered *item*. This is a familiar lesson from at least two of the themes we have already considered: the connectionist dissolution of the line dividing algorithm and data and the debate surrounding reconsolidation and its apparent pervasiveness.

Michaelian's model, like the models of many contemporary memory researchers, is clearly informed in part by connectionism. He specifically defends the trace as a distributed trace, and like many other memory scientists explicitly rejects the naïve picture of a localized, stored item in the brain.[461] The collapse of trace and system, though, just is the connectionist collapse of data and algorithm. Researchers working with artificial neural networks have many concepts and tools at their disposal for describing and predicting the behavior of these networks: none of these are, or are functionally identical to, the trace. In his 1998 book, John Sutton defended the use of the concept of "memory trace" even in a connectionist and distributed sense, first asserting that the terminology used is comparatively unimportant.

> Retention of the terms 'representation' and 'trace' is unimportant, save for polemic. But I want to show that their use in distributed models does not violate these strictures: although memory is not a notebook and 'the brain is not a writing' (Wittgenstein 1974, para 131; 1982, para 806), and although 'the word itself is somewhat disreputable' (Goldmeier 1982, ix), it is still worth retaining it.[462]

In this book, Sutton synthesizes some Wittgensteinian considerations, modern cognitive science, and historical and cultural aspects of our concepts of memory admirably. He surveys many of the same issues and difficulties we have surveyed above, and draws many similar conclusions. His insights on connectionism and the history of conceptions of memory have influenced this present project. Nonetheless, in this passage he is mistaken in two claims: first, that the concept of trace is worth retaining even after learning connectionist lessons, and second, that the language we use to describe our models is "unimportant, save for polemic." Indeed, although he has not explicitly disavowed either of these claims, Sutton's recent work treats the trace must more ambivalently, admitting that it is not clear whether the trace can survive this transformation after all.

> There is disagreement about whether these same worries also hit the mark against the radically revised notions of the internal memory trace defended in some quarters of post-connectionist philosophy of mind, mainstream cognitive psychology, and dynamical and systems neuroscience.[463]

461 Michaelian 2011a, p. 331.

462 Sutton 1998, p. 278.

463 Sutton 2014, p. 5.

Sutton goes on (in this article and in others) to remain neutral on the existence or appropriateness of a trace in most of his characterizations of memory. He is right to do so; the neural network draws no distinction between trace and system, between data and process, and neither should we.

Lastly, this collapse is familiar from debates concerning reconsolidation. As we have seen above, the reconsolidation of the memory trace has been found to be embarrassingly pervasive. First the trace was consolidated, and then it was reconsolidated, and now it seems to be undergoing continuous reconsolidation, which is corollary to the constructive activity that Michaelian's generative model is supposed to account for. As we have seen, memory scientists are worried about the ubiquity of such a process. If the trace is never *not* being changed and reconfigured, is there any sense to even positing this process of "reconsolidation"? Yes and no. The collapse of the reconsolidation "process" working on the sometimes-static "trace" is once more just the collapse of data and process. The trace-system combination is inherently dynamic, as we have learned from the behavior of connectionist systems and from the remembering behavior of humans. Every approach in which we attempt to separate out the trace from the system or process working on it or transforming it is an approach that stalls. Continuous reconsolidation is still a useful concept, but not the continuous reconsolidation of some static object that is only dynamic contingently when worked on by an appropriately dynamic process. Rather, we might say that our disposition to remember something is continuously "reconsolidated", if by this we only mean that it is constantly changing. Furthermore, it is clear that some cognitive and neural stimuli can change our performance of this memory more profoundly than others. Although the disposition to remember something is never static, it is in some conditions more dynamic than in others—these conditions already constitute the object of study by many cognitive neuroscientists.

5.2.3 *Particularities without components*

Some ways of remembering are also ways of engaging in other cognitive processes and tasks, such as perception, action, and speech. Remembering can happen in the most varied ways, and yet the very richness of this variety precludes fixed categories. This is part of the reason Brockmeier foresees memory's dissolution as a unitary object of study.[464] This preclusion of categories also seems to spell the doom of *memory systems*, which, despite a few lingering controversies, had provided one of the more successful and widely adopted developments in the cognitive psychology of remembering.[465] In the following section, I want to highlight some of the ways that varieties of remembering can be (and are already being) characterized and studied even in the absence of component memory systems or unifying component processes.

First it should be noted that studying memory in the myriad forms it takes in everyday cognitive life is no new endeavor. Although this present project has focused primarily on the study of memory in cognitive psychology, philosophy, and cognitive neuroscience, there are many other fields and methods that study human remembering. Anthropology, education, history, cultural studies, sociology and other social sciences and humanities all have a long history of studying assorted phenomena of remembering, and there has recently been some movement toward further integration among research programs in these disciplines and in psychology, philosophy, and neuroscience.[466] Indeed, one barrier to such integration is just the commitment to the computationalist assumptions above: the reason that disciplines outside of cognitive psychology and neuroscience figured little into the above critique is that these other disciplinary approaches

---

[464] Brockmeier 2010.

[465] Squire 2004; Craik 2007.

[466] Roediger and Wertsch 2008; Radstone 2008.

are considerably less prone to latent computationalism. The basic model of the storage and retrieval of traces, it has been argued, has persisted in work in cognitive psychology and neuroscience even in spite of explicit disavowals and cosmetic revisions. For many other disciplines, this simply has not been so. Social psychologists engaged in the study of narrative cultural memory or transactive memory among groups employ mixed qualitative and quantitative methods, not tending to model memory in computational terms.[467] Cognitive anthropologists studying the role of memory in natural human cognitive behavior have also tended to avoid the trappings of computationalism.[468] Social scientists and humanities scholars studying trauma, historical narratives, or memory and aging, among many others, tend to do so as well. Cognitive neuroscience and certain strands of cognitive psychology seem particularly prone to latent and problematic computationalism.[469]

There is increasing collaboration across these disciplinary lines. Perhaps in response to the transdisciplinary influence of work in cognitive anthropology,[470] cognitive psychologists are once more emphasizing real-world remembering.[471] These increased efforts to attend to the many forms that remembering takes, as well as the corresponding concerns about the simplifications imposed on studied memory by laboratory constraints and assumptions, are in accord with Wittgenstein's exhortation to attend to particular cases. Furthermore, since the research programs

---

467 Webner 1987; Brockmeier 2002.

468 Hutchins 1995, 2000; Bloch 1998.

469 Jens Brockmeier, having done work in social and cognitive psychology, expresses surprise at the extent to which cognitive psychology and neuroscience have clung to perspectives not assumed by so many other fields and neighboring disciplines (Brockmeier 2010).

470 A notable example is the work of Hutchins (1995; 2000).

471 See the collected articles in Cohen and Conway's (2007) volume *Memory in the Real World*.

outside of cognitive psychology and neuroscience already tend to study ways of remembering without assuming modularity or computational subprocesses, these transdisciplinary collaborations can only help to mitigate many of these resultant computationalist difficulties. If a couple misremembers an event from their biographies, nobody argues about whether this is a failure of transactive *or* a failure of narrative memory; these are not taken to be modular and distinct. Each of these are taken to be patterns of remembering that include many different cases, but it is accepted that they overlap with one another and have fundamentally indeterminate boundaries.

This points to a second promising area of development in the study of memory. Debates about the grounding of memory systems have been leaning toward functional bases in recent years, rather than distinct neural bases.[472] That is, although memory systems are still often treated as neurally distinct mechanical systems,[473] there are a number of researchers who advocate the characterization of "memory systems" in terms of their function or processing types rather than in terms of their underlying neurobiology. This approach to memory systems has been influenced by the work of Fergus Craik, who has been cautiously but persistently critical of the notion that systems are separate storage areas in the brain for several decades.[474] Craik's models sometimes evince a familiar computationalism, but without the emphasis on distinct brain systems that underlie distinct memory systems:

> By the present view, there are no different stores or systems, only different processing operations, representing sensory, phonological, visuospatial, conceptual or other types of

---

[472] Craik 2007; Vosgerau 2010; Henke 2010.

[473] Squire 2007; Buckner 2007.

[474] The starting point for this is an influential paper by Craik and Robert Lockhart (Craik and Lockhart 1972), though Craik's work has clustered around these criticisms in the intervening decades (Craik 1992, Craik 2007).

> information, which act to modify relevant existing representational schemes, in part by adding episodic records to appropriate representations.[475]

Other researchers have continued this program, attempting to provide an alternate basis for different ways of remembering.[476] This move toward functional descriptions of the ways that humans remember is one way to describe and study different ways of remembering without committing the fallacious move from variety to distinct systems. If episodic or procedural remembering are taken to be theoretical constructs of models, there is no reason that cases like repertory memory cannot simply exist on or near the ill-defined boundaries. Take models in ecology as an example. Ecologists divide species into two groups, depending on whether they tend to invest reproductive efforts in many distributed, uncared-for offspring ("r-strategists") or whether they tend to invest reproductive efforts in a few concentrated cared-for offspring ("K-strategists"). These are elements of robust models, defined by rigorous and fruitful equations. Yes there are boundary cases, of course—some organisms, like gingko trees and sea turtles, simply exhibit mixed traits.[477] This worries nobody in ecology: K-selection and r-selection are theoretical constructs. Certainly no ecologist would suggest that whether sea turtles belonged to one group or the other could be resolved by dissecting a turtle and seeking physiological correlates of r/K-selection. The same could be said of nearly any concept in ecological modeling.[478] The move toward functional descriptions of the ways that humans remember is just also a move toward this modeling approach.

---

475 Craik 2007, p. 132.

476 Roediger et al. 2002; Henke 2010.

477 Wilbur et al. 1974.

478 Another example might be "apex predator", which is an essential concept for many successful ecological models. Nonetheless there are many actual species that are simply borderline cases of apex predation, and this indeterminacy is accepted as part of the construct.

Perhaps as a consequence of this modeling and functional basis that is increasingly accepted for memory systems, many memory scientists have already been focusing on the ways that remembering may be neurally coextensive with a wide range of other cognitive activities. Anthony McIntosh follows the above "levels of processing" definition of memory systems to an almost explicitly Wittgensteinian conclusion:

> To restate, an important implication of the neural properties discussed here is that all processes in the brain have the capacity to lead to or modify memory. The brain seems particularly well adapted to encode and store information at all levels. It is tempting to say that memory in its highest form (e.g. episodic) is a derivative of the standard operations of the brain. While psychological theories tend to differentiate memory from nonmemory operations such as sensation and action, it is important to recognize that sensation, perception, attention and memory are intimately intertwined. The very acts of seeing, hearing and acting make use of the brain's capacity for memory. This is probably why, in typical circumstances, what forms the contents of our memory is not under intentional control. Indeed, those processes that facilitate memory (e.g. levels of processing) are essentially by-products of other cognitive operations.[479]

McIntosh is right, and this direction in memory science—toward function and theoretical models treated as such and away from subcomponents that match subsystems—signals real progress in the field. Cognitive neuroscientists who are puzzled by the "quite mysterious" biochemical processes that the models of psychologists assure us neural correlates of memory *must have* will do well to heed those psychologists who are changing their minds about the implications of these models. As Craik writes:

> This account of encoding processes has been couched in purely cognitive terms, but it has clear implications for corresponding neural activities. First, memory-encoding processes should be indistinguishable from the neural activities associated with attending, perceiving and comprehending. Secondly, the correlates of retrieval processes should overlap those occurring during encoding, in part at least.[480]

---

479 McIntosh 2007, p. 62. Note that McIntosh still uses 'encoding' and 'storage'. Brockmeier also makes this point about the overlap between perception and memory in particular, writing "there is no evident distinction between brain processes operative in remembering and in perceiving" (Brockmeier 2010, p. 20).

480 Craik 2007, p. 133.

The neural processes that support remembering are indistinguishable from (because they are identical to) the neural processes that support other cognitive capacities. Memory scientists are studying, and should continue to study, the different ways that humans remember. Neuroscientists are studying, and should continue to study, the neural processes associated with various cognitive performances. Neither of these groups have any need of the assumption that a distinct and modular way of remembering is also, or is supported by, a distinct neural process.[481]

5.3 Causality and isomorphism in memory

We have seen that the concept of a memory trace has been posited in order to account for the apparent circumstance that past events cause future remembering. This causation, it is thought, requires there to be a physical trace linking the two.[482] Furthermore, since the later remembering has features that are structurally analogous to that which is remembered, the memory trace must account for this isomorphism by containing something structurally analogous to these features. This view has been reinforced by the existence of many physical systems that really do duplicate information once received by preserving this information in the readable structure of some physical thing. Computer storage works this way, but so do magnetic tapes, vinyl records, or photographic negatives, among many other examples. With the right translation rules, one can study the physical features of the tape, record, or negative and thereby infer complex structural features of the recorded and reproduced thing.

---

481 This conceptual separation without isolated subsystems echoes William Bechtel's recent attempt to reconcile mechanistic explanation with situated cognition using "nearly decomposable" functional models as mechanisms (Bechtel 2009).

482 John Sutton makes this requirement fairly explicit, noting the seemingly common-sense claim that "the past experience itself must have been causally operative in producing (intervening) states that are in turn causally operative in producing the present recollective experience. … What's surprising about this analysis is that it suggests that built in to common sense concepts of memory is a reliance on the existence of some kind of 'memory trace' as a continuous bridge across the temporal gap, causally connecting past and present" (Sutton 2012 §1.3).

However, we have seen that this explanation of memory as an isomorphic record leads to a collection of conceptual problems. It is also persistently ill-suited to the ways that remembering actually works, as these have been established empirically. These problems have led some to deny the need for, or viability of, a memory trace, but this denial has often been construed as (and has sometimes genuinely taken the form of) a denial that remembering is causal at all.[483] Causation in memory, from the past event to the present remembering, is thought to require an isomorphic trace. This requirement has been utilized by both opponents and proponents of the trace—the former take the problems inherent to the trace to imply that memory is not causal, or is non-standardly causal, while the latter take the causality of memory to guarantee an isomorphic trace. I want to show that this common claim, that causation in memory requires an isomorphic trace, does not deserve consensus or immediate assent, so that we can lay the groundwork for accounts of memory that do support regular old causation, but that do not invite the problems above, which I have claimed are inherent to the notion of a stored memory trace.

5.3.1 *Upsetting our concepts of causality*

Philosophical arguments about the possibility, necessity, and role of an isomorphic trace have suffered in the absence of a common framework employed by the parties to this debate. Those who attack the notion of the trace and those who defend it often talk past one another. Opponents of the trace have voiced concerns about homuncularity and kindred difficulties taken to follow from the postulation of a trace,[484] and proponents of the trace have voiced concerns

---

[483] Malcolm's (1977) account, discussed in more detail below, rejects both traces and the notion that memory is a causal process. The "direct realist" account of Wilcox and Katz (1981), though couched in different language, also characterizes memory as non-causal.

[484] Malcolm 1977; Wilcox and Katz 1981; Shanon 1998; Dalla Barba 2001; Brockmeier 2010.

about causality in the absence of a medium.[485] Some of Wittgenstein's own remarks on the trace may have been responsible for exacerbating this impasse: Wittgenstein, as quoted in chapter four, takes the dubious nature of the memory trace as good reason to upset our very notion of causality. Let us more closely examine the above quoted passage from *Zettel*:

> I saw this man years ago: now I have seen him again, I recognize him, I remember his name. And why does there have to be a cause of this remembering in my nervous system? Why must something or other, whatever it may be, be stored up there in any form? Why must a trace have been left behind? Why should there not be a psychological regularity to which no physiological regularity corresponds? *If this upsets our concept of causality then it is high time it was upset*.[486]

There is a reading of this passage on which Wittgenstein is straightforwardly claiming that we can and should attribute causality even in the absence of *any physical connection* between the cause and the effect. Indeed, this reading has been accepted by some who have explicitly attended to this passage, and more importantly the belief that a denial of the trace necessitates causality-at-a-distance has been instrumental among defenders of the trace. To the physicalist, or even simply to the researcher with vaguely physicalist inclinations, it can seem positively anathema that any phenomenon should be caused by some other phenomenon to which it has *no physical connection*.

I will argue for another reading of this passage. I think there are good reasons to believe that my alternative reading is one that would have been countenanced by Wittgenstein, some of which I will sketch, but my primary aim is to read this passage in a way that resolves the impasse between proponents and opponents of traces, irrespective of authorial intent. There is a way forward, between causality-at-a-distance and isomorphic traces, which is gestured at in this

485 Rosen 1975; Sutton 1998; Bernecker 2001;

486 *Zettel*, §610, added emphasis.

passage and other related passages. In order to see this, let us consider what could be meant by a "cause of this remembering in my nervous system."

In the causation-at-a-distance reading, the supposition that there is a cause of this remembering is taken to mean that there are causal connections between the two states: between the total physical state of someone who has not yet remembered and the total physical state of that same person once she has remembered. Either the later "remembering" state is, at bottom, a physical state or not. If it is, then to deny that there are any causal connections between the two is to claim that the later state is either utterly causeless or that causality need not supervene on the physical. If the remembering state is not physical, causality is still not physically mediated, and, perhaps worse, there may be spooky non-physical entities involved in this explanation. Either way, it appears, Wittgenstein is making a wild and idiosyncratic claim about the nature of causality and the physical world. It is no surprise, on this reading, that he ends the passage exhorting us to upend our basic notions of causality. Memory researchers who defend the trace against this possibility, even philosophers, sometimes admit that they are imposing an empirical claim on memory ahead of study, albeit a weak and eminently plausible one. John Sutton sums this up nicely:

> The postulation of traces *is* empirical, but the relevant domain is not psychology. What's doing the work is the *physical* assumption that there is no macroscopic action at a temporal distance, that mechanisms in fact underlie apparent cases of direct action between temporally remote events. This assumption may be mistaken, but challenges to it must offer some positive alternative theoretical framework.[487]

However, there is another reading of what might be meant by the supposition that there is "a cause of this remembering in my nervous system," and this reading may offer just the alternative framework Sutton is requesting. We might alternatively take this to mean *that there is some*

[487] Sutton 2012, original emphasis.

individuate-able thing, in the nervous system, which is identifiable in principle as a cause of the later instance of remembering. To deny this claim is merely to deny that any one thing, configuration, or structure can be individuated, even in principle, as the cause of the later behavior. This is not to deny that there are any causal connections between the total states before and after remembering, but only to deny that any part or subset of features of the first state can be picked out as causal for the later event or process. If this reading of the phrase is correct, it makes sense that Wittgenstein leans on psychological and physiological *regularity* to support it: the denial of a second-reading "cause of this" in my nervous system is the denial of meaningful mappings between any *patterns* in the remembering process and any *patterns* in the stored trace. The nervous system surely has some regularities, as the remembering behavior also does, but we will not find any correspondence between these.

In the paragraphs that lead up to this passage in *Zettel*, Wittgenstein constructs an example of this kind of lack of correspondence.

> No supposition seems to me more natural than that there is no process in the brain correlated with associating or with thinking; so that it would be impossible to read off thought-processes from brain-processes. I mean this: if I talk or write there is, I assume, a system of impulses going out from my brain and correlated with my spoken or written thoughts. But why should the *system* continue further in the direction of the centre? Why should this order not proceed, so to speak, out of chaos? The case would be like the following—certain kinds of plants multiply by seed, so that a seed always produces a plant of the same kind as that from which it was produced—but *nothing* in the seed corresponds to the plant which comes from it; so that it is impossible to infer the properties or structure of the plant from those of the seed that it comes out of—this can only be done from the *history* of the seed. So an organism might come into being even out of something quite amorphous, as it were causelessly; and there is no reason why this should not really hold for our thoughts, and hence for our talking and writing.[488]

It is worth noting that Wittgenstein's example maintains a physical connection across the causality in question—something it would not have to do on the first reading of what he means by the denial

488 *Zettel*, §608, original emphasis

of a "cause of this in the nervous system." Rhetorically, however, I fear that the details of Wittgenstein's example are somewhat unfortunate. Perhaps contingently, seeds do have internal structure, and there are regularities in this structure that at least appear to have mappings, available in principle, to the phenotypic structure of the plant they will become.[489] That this is even arguable makes the example suboptimal. Alternatively, Wittgenstein may just be asking us to *imagine* a seed without such structure—he often and fruitfully engages in suppositions that upend all manner of natural facts—but if this is the case then while the story he tells may elucidate his claim for those who are already convinced, it is unlikely to convince his opponents. Anyone who was already skeptical about this kind of causality-at-a-distance will be no more receptive to it in the form of a seed than they are in form of a memory. Perhaps this has contributed to the impasse as we have found it between those few who have been convinced that there need be no trace, and the many who remain unconvinced.

In *Memory and Mind*, Norman Malcolm offers an argument with some parallels to the one I offer here. Malcolm, however, draws exclusively on examples of processes that are, themselves, cognitive—a tactic that is unlikely to convince those who are trying to understand how best to understand cognition, and who remain committed to the contiguity of isomorphism as a requirement for the contiguity of cause.[490] Furthermore, Malcolm ends by determining that remembering cannot be accurately described to be a causal process at all. Malcolm's argument hinges on the fact that there cannot be said to be any ultimate "structure" to an experience at all. While this claim is one with which I am sympathetic, and I take it that it is one of a family of

---

489 Although Varela and colleagues make a persuasive case that the story is not so simple even in the case of the genetic "code", emphasizing that this can only be understood as a "code" with "translation rules" if a vast tract of context is taken for granted (Varela et al. 1991).

490 Malcolm 1977, p. especially p. 236-239.

claims that has been used successfully in many other instances,[491] it does not support Malcolm's conclusion. We do not need to identify the "ultimate", ontological structure of an experience in order to show that a remembered event, or experience, has structural similarity to some later process of remembering. It clearly true that there is an identifiable isomorphism between *some* structure found in the portrait I draw of someone I remember and *some* structure in the past acquaintance's actual visage. Does this similarity get at the "ultimate" structure in either of these? Perhaps not, but this matters little for the structural analog we may or may not be seeking in the memory trace that mediates between the two. If we listen to a song and then some moments later produce it by whistling, there is certainly identifiable structure shared between these processes, irrespective of said structure's ontological ultimacy. The question is whether this structure, whatever its ontological status, is also shared by the memory trace. Most researchers, disregarding the sentiment in the stronger reading of Wittgenstein above, claim that memory is a causal process and so there *must be* some structural analogue of the whistled tune in the latent trace. Malcolm, conversely, heeds the strong reading of Wittgenstein and ends up denying that remembering is rightly called a causal process at all.[492]

My aim in the following considerations is to take the milder reading of Wittgenstein as a way between these. In an important sense, remembering is a causal process. Nonetheless, there

---

[491] This seems to be a corollary of Wittgenstein's general move away from atomism to holism, and the claim that nothing has the meaning that it does apart from its context in the whole "flow of life" (Stern 1995). Similar points are used by Dreyfus to argue that programming a machine to recognize "relevant" features of a situation faces deep and serious difficulties (Dreyfus 1972/1992). Boncompagni writes, "In a nutshell, at the core of Wittgensteinian cognitive science is the inextricable connection between mind and action, knowledge and practice, language and life, nature and culture" (Boncompagni 2013, p. 30). That I am not focusing on this Wittgensteinian lesson here does not mean that it has no bearing on present memory science—Gallistel and King characterize the trace such that it is isomorphic to the actual "formal structure of the world" (Gallistel and King 2011, Ch. 4).

[492] Malcolm 1977, Ch. 7. Malcolm does deny "causal process" status to remembering, but he characterizes many of the phenomena of remembering in causal terms, and he does not straightforwardly deny the need for a physical intermediary.

need be no cause in the physical medium of this remembering, which persists across the temporal gap between remembering and that which is remembered. Despite being causal, there need be no physiological regularities that correspond to the regularities we find in both the earlier and later phenomena. An account of memory that succeeds in avoiding the problems of the trace, but also fits into what we do know about mind and brain, will need to be in terms of just such a causal, *intermittently* isomorphic, process. In order to show that this is unproblematic we need examples of processes that fit this profile. For those already convinced, examples abound in the way that the past behavior of cognitive systems influences later behavior.[493] Additionally, artificial neural networks can be interpreted as supporting this kind of causation with unstructured mediation—but again, the appropriate characterization of neural networks remains controversial. If our aim is to convince the defender of the trace that traceless remembering need not invoke nonphysical entities or spooky causation-at-a-distance, there are better examples available.

5.3.2 *Two wobbles and a lump*

Let us clarify the elements needed for this example. The supposition is that there can be two processes, call them **α** and **γ**, which take place at t_1 and t_3, respectively. At t_2, neither **α** nor **γ** are taking place, but the physical substrate **β** persists through t_1, t_2, and t_3. Additionally, there are regularities in **α** that correspond to regularities in **γ**, but there are no regularities in **β** that correspond to regularities in **γ** (and, presumably by transitivity, nor are there regularities in **β** that correspond to regularities in **α**). In this abstract form, on my preferred reading, **β** can play a causal role in **γ**, just as long as there are not isomorphic relationships between features of **β** and features of **γ** such

---

[493] Indeed, Malcolm and others who have argued against the need for a trace (e.g., Wilcox and Katz 1981; Putnam 1999, p. 132; Randall 2007) tend to use psychological phenomena as examples of cause with unstructured mediation. It is perhaps unsurprising that these have been dialectically unsuccessful, since the causal explanation of psychological regularities is what is at issue.

that there is any part or thing in **β** that we can individuate as a cause of **γ**. Furthermore, an example will be best in which **α** can, more or less unequivocally, be said to be a cause of **γ**, despite the fact that there is no such cause to be found in **β**. It is important (though perhaps not ineliminable) for our purposes that **α** and **γ** are processes, but **β** is a physical entity or arrangement of physical entities. Examples of this configuration of processes and substrates are neither far-fetched nor particularly hard to come by, and examining some of these closely will help to distinguish between these two readings of Wittgenstein's denial of "a cause of this remembering in my nervous system."

Consider a lump of clay on a potter's wheel. As the potter is shaping this clay, let us imagine that the clay starts to wobble with the application of pressure from the potter's hand. It wobbles thus for a little while, and this wobble is self-reinforcing, as wobbles in spinning clay tend to be. Further, let us imagine that this wobbling process has certain features—a particular frequency, a pace at which it is extending the lump of clay in certain directions, and so on. The potter then stops the wheel, perhaps with the intention of reshaping the clay, but she gets distracted for a little while. For this duration, the wheel and lump of clay sit motionless. The potter then returns to the wheel, starts it spinning, and applies similar pressure by hand. The clay, once more, starts to wobble. This second wobbling process shares many features with the first—the frequency, the rate of extension, etc., are all what they are *because the first wobble reinforced those characteristics*. Despite being asymmetric and crucially involved in both wobbling processes, the lump itself is relatively featureless.

This is a case of two processes, the initial wobbling and the revived wobbling, which are causally mediated by a thing, the lump of clay. The two processes do not overlap in time, and it is clear that there are regularities in the second process that correspond to regularities in the first process. There is at least one, fairly robust, sense in which the first process can be said to be the cause of the second process. The only point that remains is whether there is a cause, *in the inert lump of clay,* of the second wobbling process.

The strangeness of this question bring out precisely the ambiguity that the two above readings trade on. In one sense, everything about the lump of clay is certainly causally connected to the clay's behavior during the second wobble. One answer to the question of why the second wobble is what it is, is simply to say that the entire composition and structure of the lump of clay makes it so, given the conditions of the wheel and the pressure of the potter's hand. In this sense, there may not be, strictly speaking, a cause *in* the lump of clay, but it does seem that the lump of clay as a whole is one candidate for the cause of the wobble.

But there is another sense in which picking out a cause in the lump of clay seems exactly wrong. This is especially so if we focus on the regularities exhibited by the second process. Imagine that, for whatever reason, an observer were surprised to see the clay's wobble so quickly and readily adopt that particular frequency. This observer asks: "Why did this process have this particular feature, why did the vibration so quickly assume that frequency?" In response, the "explanation" that the particular frequency was "because of" the whole shape of the clay in conjunction with the conditions of wheel and potter's hand seems ludicrous. Of course the whole physical system is in some sense the proximal cause of its state, but this is a trivial—one might even say nonsensical—sense of *cause*. Almost every phenomenon is in some sense causally dependent on the whole material state of its system. This is not what we are seeking when we seek causes. An astronomer who explained that a star went supernova "because of" the star's whole physical state in conjunction with its conditions, or a botanist who explains that a flower blooms at night "because of" its whole physical state and its conditions, would be making a mockery of causation. When we individuate causes, we pick out relevant and corresponding features or regularities. In this sense, the cause of the second wobble's ready frequency is that the system underwent a similar and self-reinforcing process earlier. The regularities in the first process correspond with, and cause, the regularities in the second process, but there are no corresponding regularities in the physical substrate that mediated these—even though the intervening physical substrate was a necessary condition for the second process.

It is especially helpful to contrast the clay example with the example of substrate mediation of disparate processes we met in chapter two: that of the carving and playing back of a vinyl record. In the case of the record, there are parallel causes available, depending on the context of explanation. If I want to know what caused a note in a song to sound just so, I may be asking about the musician's choices when he or she recorded the piece into the gramophone's horn, or I may be asking about how the quality of the sound has been affected by the shape of the grooves on the record and the playback system itself. This is an interesting difference, but I only need clarify which one it is that I seek in order to know how to go about finding the answer. The second possibility—that I may be wondering about the cause of the note's sound with respect to the actual grooves on the record—is viable because *there is something that corresponds to the note*, on the physical, vinyl record. It is not just the record as a whole, but very particular features that can be individuated, that map isomorphically to features in the way the song sounds.

This, I argue, is what the clay lacks. It may be objected that there is a match for the frequency of the vibration in the physical shape of the clay, but this is implausible. The lump of clay, remember, may be very simple. It has few features, and the total shape it has could have come about in many different ways. Furthermore, the lump of clay would produce many different wobbles under many different conditions. It is only such that it wobbles *thus* when the wheel is spinning just so and the potter is applying pressure just so. To say that there are "regularities" in the lump of clay that correspond to "regularities" in this wobble is simply to misuse the language involved. The lump of clay is, in an important sense, not a record. The only sense in which it is a record of what has gone on before is the sense in which *everything* is *always* a "record of" what that thing has previously undergone. The only sense in which it is or has a map of its future processes under certain conditions is the sense in which every physical thing always is or has a "map" of its future processes under certain conditions. To use the concepts of corresponding regularities in this way is just to abuse the language of records, maps, and isomorphisms.

If there are no regularities in the clay that correspond to regularities in the later wobbling, then we have found our **α**, **β**, and **γ**. The first wobble is **α**, the second wobble is **γ**, and the inert lump of clay between them is **β**. While it's true that regularities in **α** correspond to regularities in **γ**, it's not true that regularities in **β** correspond to regularities in **γ**, even though **β**, and even certain properties of **β**, are necessary conditions for the features of **γ**. The first process causes the second process, even though there is no cause of features of the second process that can be individuated *in* the intermediary substrate. The question "what is the cause of this particular frequency *in* the clay?" should be rejected.

This circumstance is not a product of any unique qualities of the objects or processes in the example. It is true that examples of self-reinforcing processes that can be revived by retention in a physical medium require both plasticity and perdurance of the medium. Clay is, obviously, a good candidate by virtue of exemplifying both of these properties, but we may just as well have considered examples such as a violin that more readily produces certain notes given its past performances, or a door that readily slams shut because it has so often in the past been slammed shut. Notably, neural networks also demonstrate this combination of plasticity and perdurance—

two properties between which there is often tension—but once more, given that the right characterization of the phenomena of neural networks is part of what we are seeking, to use neural networks as an *example* of this kind of causation would be overly hasty.

We have met with interdependent phenomena that tend toward dynamic equilibria before: these are the primary subjects of dynamical systems research. In this example, the clay, the wheel, and the potter's hands comprise a dynamical system that could be modeled using coupled equations. The complex and repeating behaviors of the system that are self-reinforcing mark what dynamical systems theorists call *limit cycle attractors*. Any complex system that tends toward certain states has attractors—these can be represented as points or regions in the system's total state-space toward which the system tends from nearby points in the state-space. Some of these attractors—like a die landing on its face rather than on its edge—are simple states of the system. Others—like the orbit of a satellite—are oscillating cycles. This latter variety of attractor is called a limit cycle. When a complex system is exhibiting limit cycle behavior, it will tend to return to this behavior even after displacement.[494] In the case of the clay, the wheel coming to a stop is just the transition to another point in the system's state-space (for which the shape of the clay has remained constant, but the velocities of the components have been changed to zero).[495] The resumption of the wheel's spin and hand's pressure moves the system to a point near to the original attractor cycle, which results in a return to the same limit cycle behavior.

Dynamical systems researchers have assembled an extensive toolkit of resources for characterizing, modeling, and predicting the behaviors of systems like this one. In dynamical

---

[494] Gregor Schöner writes, "Behavioral patterns resist change; that is, they are stable. This may be mathematically characterized by considering behavioral patterns as the attractor states of a dynamical system" Schöner 2009, p. 25).

[495] Technically the wobble is a *local* attractor, since there are points in the system's state-space that do not tend toward the limit cycle in question (take, for example, the flattening of the clay or the death of the potter).

systems terms, the initial shaping of the clay just was the moving of the system into states that were nearer to, and gradually became more steeply inclined toward, the limit cycle attractor that the wobble constituted. The second instantiation of this attractor—the second wobble—is unsurprising: this tendency to return is what defines an attractor. The time interval between attractor states, without any "record" of those repeated features of the attractor in the interim, is of no consequence. Causal claims do not play any special role in the dynamical explanation of systems and behaviors like these: the attractor cycle at t_1 naturally shares features with the attractor cycle at t_3, as this is just what it is to be an attractor. If we ask how the system came to be in the state that tends toward this attractor, the answer is just that it had already moved toward and into the attractor from a series of nearby states. The first wobble provides an explanation of how the second wobble came to be, and the features of the first wobble map nicely onto the features of the second wobble. The interim states were necessary for this continuation, but any search for matching features in the interim state is misguided. "The" cause of the second cycle is not to be found in the interim state.

So when Wittgenstein urges us to upset our notions of causality, let us read this exhortation in a slightly less revisionary manner. This reading does not deny that there are, in a somewhat trivial sense, causal links between the physical substrate and the resulting process. Rather, it denies that there are any regularities in the physical medium that correspond to the regularities in either process. This reading emphasizes that to seek a cause for this *particular* feature of some process requires a corresponding regularity. We have not found a cause for this behavior in the clay if there are no regularities that can be identified that correspond to the behavior in question. If the fact that **α** causes **γ** even though there is no individuated cause of **γ** available *in* **β** means that this causation is happening "at a distance," so be it. This distance may upset our notions of causality, but it does not upset them quite as much as they would be upset by the first-reading version of processes and causation as sketched by Wittgenstein.[496]

Let us extend the metaphor of the potter's clay, just a bit, to demonstrate how this set of processes and substrates can illuminate our assumptions about memory traces. There is no structural analogue of the first wobble's frequency identifiable *in* the clay. The total shape and structure of the clay lump is a product of the first wobble, among other things, but that is all that can be said. The frequency of the first wobble was no necessary condition for the structure of the lump—for the lump could have taken that shape in any number of ways—and neither is the frequency alone a sufficient condition for the structure of the lump.[497] Everything about the first process, the initial state of the clay, the speed of the wheel, the pressure and shape of the potter's hand—perhaps even the Coriolis effect of the rotating earth—all taken together, provide jointly sufficient conditions for the structure of the resulting lump, but only *all* of these conditions taken together provide such. Similarly, the mere shape of the lump is not a sufficient condition for the second wobble's frequency. The frequency of the second wobble cannot be "read off" of the shape of the lump. The lump of clay will exhibit all sorts of different motions and vibrations, depending on exactly what it undergoes. There are no "translation rules" to translate the lump's structure into a frequency, unless *all* of the other conditions of the second process are specified. That is, the only translation scheme involves filling in all of the details that make up the second process. There are

---

496 This procession of cognitive features out of the non-featured chaos of the brain might be reminiscent of Searle's insistence that certain aspects of cognition are biological, brain-dependent products (e.g., Searle 2007), which has been criticized for its lack of explanatory heft (Hutto and Myin 2013, p. 143). Although Searle's explananda are not my explananda, something like the above account could help to divide what needs explaining from what does not, to lend some explanatory heft to a project like Searle's.

497 Again the state-space characterization is helpful. The clay's shape and motionlessness at t_2 comprise one point in the system's total available state-space. This point is no record of the past: unless the state-space has very specific (and implausible, in this case) topologies, any number of tracks through its state-space could have brought it to this point.

no *rules* in this translation, only *particular circumstances*. Is there any sense in which the second wobble's frequency already exists in an encoded form in the clay, before the process begins? Only in a trivial and misleading sense, since by admitting this state of affairs as an encoded event must also mean that the same "encoded" structure—the whole of the lump of clay—contains the stored information required for an infinite number of processes to which it would contribute under various conditions. If this is what encoding and "information storage" are, then everything everywhere is encoding almost everything else, all of the time.[498]

If there can be such processes, then it is a mistake to assume that causation between two processes whose respective structures bear similarities to one another implies that the shared structure must also be found in the physical substrate that mediates the temporal gap between these processes. There can be causal links between processes with isomorphic structures, even though the causation is mediated by something that does not exhibit this isomorphism. Given this, whether the resemblance between the remembered event and the remembrance is mediated by a neural trace *that also shares identifiable structure with these events* is an open question. Some causal processes do work this way—witness the carving and playing back of a vinyl record—and some do not. Once it is established that the latter is a possibility, it is not very difficult to find some evidence that memory fits the traceless-causation framework better than it fits the trace-causation framework, especially given the problems produced by an isomorphic trace noted above.

---

498 This argument has obvious cousins in various indeterminacy worries across several philosophical domains. One is "Searle's Wall", the difficulty with certain descriptions of computationalism that, according to Searle, imply that, for example, the ambient molecular behavior of the paint on the wall can be functionally mapped to, and hence interpreted as, a word-processing program, given the gap between syntax and semantics (Searle 1993). Notably, Chalmers responds to Searle's concern by emphasizing the external, behavioral, interactive features required for genuine computation or cognition (Chalmers 1996). A parallel problem is raised, for causation in particular, by Ned Block (Block 1990), who worries that mere syntax is too indeterminate to provide the causation we find in folk psychological states. Carrie Figdor responds to what she calls "Block's Paradox", also emphasizing extrinsic features, by relaxing the requirements of computationalism to include computational transformations whose sensitivities surpass mere syntax (Figdor 2009).

Indeed, Wittgenstein is not alone in speculating about the ways in which the phenomena of memory may demonstrate particular forms of causation. In *The Analysis of Mind*, Russell considers a type of causation very similar to this, but which is idiosyncratic to memory processes. Russell notes that the past seems to directly influence the present state of organisms in a wide variety of cases. These past influences do not merely precede present effects in a *chain* of causes, Russell explicitly notes, but rather they (at least seem to) directly affect the present state of the organism. Russell argues that this "mnemic causation" is at least warranted as a provisional theoretical construct, and although he claims that it will likely not be found to be "ultimate" (that is, a chain of causes will later be revealed for each of these phenomena), he very clearly claims that there is nothing impossible about this style of causation:

> I do not think we ought …to reject entirely the possibility that mnemic causation may be the ultimate explanation of mnemic phenomena. I say this, not because I think it likely that mnemic causation is ultimate, but merely because I think it possible, and because it often turns out important to the progress of science to remember hypotheses which have previously seemed improbable.[499]

It is not always clear whether the "mnemic causation" Russell is characterizing just amounts to causation at a temporal distance. He contrasts this causation type with everyday proximal causation's requirement that cause and effect are contiguous in time, but he also uses only examples in which the causation is physically mediated across this temporal gap. The ambiguity in the concept of a "causal chain" is analogous to the ambiguity in Wittgenstein's phrase 'a cause in my nervous system'. If 'causal chain' means a chain of individuated causes with matching regularities, then Russell's mnemic causation defies just such a chain. However, if 'causal chain'

---

[499] Russell 1921, p. 92.

means a causal influence transmitted through a physical substrate from the cause to the effect, Russell's mnemic causation can still be said to take part in a causal chain.

So once more, there are two interpretations available, one of which is a more radical adjustment to common notions of causation than the other. Russell's own characterization seems to be teetering between the two, and critics have tended to read Russell's claim in its more alarming guise.[500] In order to defend the need for a memory trace, Sven Bernecker accepts Russell's challenge and sets out to demonstrate that mnemic causation is not plausible after all.[501] After emphasizing the temporal gap in Russell's mnemic causation, Bernecker writes that the problem is that mnemic causation "cannot explain where and how information is stored within the memory process." Bernecker makes it clear, though, that the awkward and counterintuitive consequences of this kind of causation follow directly from the assumption that memory is a matter of information storage. He claims that the mnemic causation account can either point to intermediary causal features as an explanation—in which case regular garden-variety causation is really what is operative—or the reproduction of information cannot be explained:

> But if the cognitive process underlying trace-free retention of information cannot be explained, memory becomes a magical process bearing some resemblance to telepathy and clairvoyance. Thus, the account of information storage based on the trace theory is not only more convincing than the account of information storage based on mnemic causation but also, when spelled out, the latter collapses into the former. *The very idea of memory storage calls for the stipulation of traces*.[502]

---

500 C.D. Broad characterized Russell's position in the less radical fashion thus: "Although there is continuity between the *total* cause and the effect…yet there is no continuity between the effect and *each* independently necessary factor in the cause" (Broad 1935, p. 459 as quotes in Bernecker 2008, p. 42). Broad still rejects Russell's mnemic causation by virtue of the non-existent causes it necessarily invokes—a criticism that Bernecker undermines (Bernecker 2008, p. 42).

501 Bernecker 2008, §3.4

502 *Ibid*., p. 46, added emphasis.

In this last claim, Bernecker and I agree. Furthermore, it is not clear that Russell's account of memory in *The Analysis of Mind* avoids the trappings of information storage (although Russell never explicitly characterizes memory in these terms)—Bernecker could be right that Russell's stronger version of mnemic causation is insufficient within Russell's framework for understanding memory. More importantly, Bernecker is right to insist that an account of information storage will necessarily struggle in the absence of a trace. In an important sense, *stored* information just is a trace. Of course, this implication cuts both ways. Bernecker takes it to mean that the success of the archival account of memory ensures the success of some kind of trace theory, but in the considerations above I have made the case that the failures of the trace imply failures for an account of memory in terms of stored information.

However, the above examples show that Russell's characterization of a possible "mnemic causation" makes it sound too idiosyncratic. Perhaps this is part of the reason that Russell thought it was improbable that this causation was "ultimate". Processes that have regularities that are caused by earlier regularities despite a temporal gap between *corresponding regularities* are not limited to psychological, or even biological, phenomena. Processes like the wobbling lump of clay on the potter's wheel exhibit sufficient complexity of features and self-reinforcement to be mediated across such a gap by a physical substrate that lacks any features that correspond to these regularities. Many dynamical systems fit this description, and these are studied and characterized in ways that do not invoke any magical processes akin to clairvoyance or telepathy. While complex and self-reinforcing processes that are prone to gaps mediated by holistic, plastic, physical substrates are perhaps not overly common, they are also not unique to ceramics and remembering. Remember the puzzlement expressed above by neuroscientists who are trying to understand what "quite mysterious" molecular mechanism could possibly start a chain reaction, and then restart it again after a halt.[503] This is exactly the (relatively non-mysterious) situation described by any dynamical system with limit cycle attractors, and it is exactly the non-mysterious situation we encounter on the potter's wheel. These circumstances only become mysterious as a result of unwarranted assumptions about the neural correlates of remembering.

5.4 Remembering without storing

The collection of phenomena that comprise remembering are resistant to models based on encoding, storage, and retrieval, and this resistance is becoming increasingly clear. In response to this, Jens Brockmeier asks, "Now, with the old notion of the archive dismantling, is there a new notion of memory in the offing that has the potential to replace it?"[504] Brockmeier answers this question in the negative, doubting that memory phenomena exhibit sufficient robustness and unity to be the proper object of this kind of scientific study.[505] Yadin Dudai and colleagues, on the other hand, also recognize that the phenomena of memory have been increasingly revealed to fit poorly with the conceptual frameworks brought to bear on them, but their response is to address, clarify, and to some extent revise these concepts.[506] Many cognitive psychologists of memory share Dudai's optimism, but it was shown above that the resultant revisions may have fallen short of being able to successfully free models of memory from the difficulties that archival conceptions have wrought. What I hope to have shown is that we can and should strike a path between these

---

[503] Sweatt 2007, p. 212, as quoted above in §2.3.3.

[504] Brockmeier 2010, p. 26.

[505] As we have seen, there are precursors to Brockmeier's pessimism from Deutscher (1989).

[506] Dudai and colleagues have edited an entire volume of short essays devoted to just this (Dudai et al. 2007), as evinced above in section 1.1.

responses.[507] It may be that memory can be studied, modeled, and characterized in the absence of storage and retrieval, but the necessary conceptual revisions must run deep.

Much of this work is already being done, and will undoubtedly continue. Cognitive psychologists like William Randall have been recently making persuasive cases for abandoning computer metaphors.[508] Dudai and colleagues are representative of many cognitive psychologists who have also begun entirely eschewing the word ‘storage’. An increasing number of cognitive neuroscientists are beginning to doubt whether any particular brain region or system is a “correlate” of various types of remembering, or of remembering in general.[509] Researchers in social psychology and other areas of memory studies have long stopped characterizing memory in terms of archives or storage.[510] Cognitive psychologists and neuroscientists debating the nature of consolidation and reconsolidation are gradually coming to consensus that reconsolidation is nearly ubiquitous. Memory scientists studying impairment are beginning to doubt whether the decomposition of remembering into successive subprocesses is explanatorily helpful. In general, the constructive, dynamic, non-archival nature of memory is on the rise, and characterizations of memory that depend on storage and retrieval are in decline. Nonetheless, progress along this trajectory has been impeded by latent assumptions about the mechanisms of remembering. This progress will continue to be stilted until these assumptions are clarified and alternative models of memory that fully reject them are on offer.

---

[507] Albeit in somewhat different terms, Carretero and Solkoff (2012) evince similar mixed optimism when reviewing Brockmeier’s claims.

[508] Randall 2007. See also Sutton 1998, Shanon 1998, and Burton 2008.

[509] Mark D’Esposito comes to this explicit conclusion in the case of working memory (D’Esposito 2007, p. 769); Jon Gabrieli surmises that there are no dedicated “memory systems” in the brain (Gabrieli et al. 1998). This marks, in some respects, a return to the pre-cognitivism anti-localization approach exemplified by Karl Lashley (1929).

[510] These are reviewed and outlined in some detail by both Brockmeier (2010) and Sutton (2014).

Fortunately, we have already met with what alternative, non-archival accounts of memory might look like, and with several ways in which these differ from standard accounts. First, genuinely constructive models such as Michaelian's generative memory are very close to avoiding the philosophical and empirical problems that traditional accounts have accrued, but these generative models cannot do away with the strict separation of "true" and "false" remembering without also doing away with the notion of a trace that can be individuated. A truly generative model of memory will avoid even a "distributed" trace, as well as any strict separation of process and product. Michaelian is right to emphasize the non-contingent nature of memory construction and reconstruction—memory is essentially constructive, not merely preservative—and a successful account of memory will captures these fundamental qualities. There is no sharp separation between remembering *x* and seeming to remember *x*—just as Michaelian insists—even though there are noncontroversial cases of each. Whether some psychological process is an instance of genuine remembering is not something that can be decided by studying the brain processes of the putative rememberer. As Michaelian's generative account suggests, when someone's recollection of the past doesn't go "too far beyond" what actually happened in the past, she has remembered. When it does go too far beyond what actually happened, she has only seemed to remember. Proust writes, "Remembrance of things past is not necessarily remembrance of things as they were."[511]

There are also no sharp separations between different ways of remembering, although there are cases of remembering that non-controversially belong to useful classifications. This view is also gradually gaining visibility, especially among researching in the "levels of processing" tradition initiated by Fergus Craik. Whether an actor who, on stage, responds to a familiar query

---

511 Proust 1925/2003, p. 131.

with a familiar answer is engaged in episodic, semantic, *or* procedural remembering is not something that can be decided by studying his brain processes. Nonetheless, some ways of remembering can be usefully modeled under a category: that densely amnesic patients in many cases exhibit dramatically impaired episodic memory but considerably less impaired procedural or semantic memory is one of the positive results of the study of impairment. Many cases of remembering will fall relatively neatly into one of these categories, and some will not.

Once more, a common feature in these revised explanations is that the models of remembering are treated as models, and the elements of these models are treated as elements of models. Accepting genuinely borderline cases between "memory systems" is only difficult when in the grips of computationalist assumptions concerning the mechanisms of memory.[512] Borderline cases between classifications in explanatory models are not rare in other research domains, as we have seen. Useful theoretical constructs in ecology or economics often apply to particular cases only partially, or in degrees. There is no more a yet-to-be-filled-in gap between hippocampal LTP and declarative remembering than there is a yet-to-be-filled-in gap between a wolf's cellular structure and the ecological concept of an apex predator, or between the actual material that runs between the north and south poles of the earth and the concept of an axis of rotation. Again, the functional descriptions of memory systems found in the "levels of processing" literature make a good start in this direction: functional descriptions tend to circumscribe cases only with indefinite boundaries.

Memory researchers in many disciplines already nominally accept that the neural conditions of remembering are properties of a neural network. These researchers also accept,

---

[512] Toth and Hunt's (1999) critique of memory systems is along these lines; they argue that memory systems can only succeed as hypothetical entities, and that they are hence unfit for the treatment they are often given in memory science.

nominally, that the cognitive system is a dynamic system. Memory science can only benefit by paying closer attention to the way that these networks and systems, as such, are characterized. Neural networks are capable of producing complex output that matches past input to differing degrees. These networks tend to exhibit the same constructive qualities found in human memory. Neural network researchers, however, do not divide process from product. They rarely speak of any inner processes at all, and questions about whether any internal arrangement or state of the network can reasonably be said to *represent* the past remain unresolved. Neural network researchers, by and large, get along with their modeling techniques without invoking concepts of encoding or storage. Notably, explanations of the abilities of neural networks do not encounter difficulties with homuncularity or infinite regress.

Similarly, dynamical systems offer a framework by which complex systems can engage in behaviors that reinforce similar future behaviors even across temporal gaps. Models of dynamical systems do not tend to invoke causation in any theoretically crucial fashion; dynamical techniques focus on describing how the system would evolve in time under different conditions.[513] As noted above, these models may not satisfy the desire for an underlying explanation of the type that Cummins demands for psychological "effects", but researchers working with dynamical models do not notice that anything is missing. In this the dynamical systems outlook is a natural ally to the Wittgensteinian outlook; dynamical systems approaches focus on detailed and elaborate description rather than explanation—or, at the very least, whatever explanation they offer is no more than what explanation can be gotten at by faithful and complete description. Although a few

---

[513] Schöner summarizes this approach to learning thus: "Cognitive processes emerge from instabilities of dynamic fields…Learning occurs as changes in behavioral or field dynamics that shift the behavioral and environmental context in which these instabilities occur" (Schöner 2009, p. 25).

researchers have attempted thorough integrations of dynamical systems approaches and frameworks from cognitive psychology, these are still quite preliminary. Much work remains.

To remember is to engage in some experience or behavior that is caused by an isomorphic experience or behavior in the past. The necessary match between past and present may have no sharp threshold. There is a causal link between past and present, but there need not be (and often is not) any interim preservation of the features of the past experience, over and above the capacity to reproduce said features under certain conditions. It may be that remembering is causal, but not in a metaphysically charged sense. Complex neural conditions must be met in order to retain capacities to reproduce features of past experiences. Characterizing the conditions—neural, psychological, and environmental—under which particular cases of remembering succeed and fail (and how they proceed) is a vast enterprise that is already well underway, and these characterizations will continue to be aided by explanatory models. The sooner this enterprise is actually freed from the latent computationalism that binds it, the quicker and clearer its progress can be.

CHAPTER VI

CONCLUSIONS

> The only real voyage of discovery consists not in seeing new landscapes, but in having new eyes.
>
> Marcel Proust[514]

The study of memory can and should proceed without invoking inner subprocesses of storage and retrieval, without invoking the notion of a stored, isomorphic trace in the brain, and without seeking to correlate inner cognitive processes with neural processes. In short, the study of memory can and should proceed without assuming that human memory is functionally analogous to what we now call "memory" in a computer. There are several ways forward on this path, many of which are already being explored. In this brief final chapter, I will consider some objections and responses, some implications that this view of memory holds for other areas of research, and some of the future avenues of research that this revision opens up.

6.1 Three objections

The following objections are by no means exhaustive, but each is indicative of a family of concerns that some of these revisions to memory science have elicited. The first is an objection from the successes of cognitive neuroscience; our understanding of the brain is sometimes already taken to have demonstrated the kind of localization I have spoken against, and this concern needs addressing. The second from language use and its implications for theory; especially outside the science of memory, it can seem as though the words and metaphors we choose cannot be so very

---

514 *The Captive*, p. 131.

damaging. Again, this is a worry worth addressing. The last is an objection concerning the nature of inner processes; Wittgensteinian characterizations of mental phenomena are sometimes accused of being overly behavioristic, and it is important to show why this account of memory does not, at least easily, succumb to classic worries about behaviorism.

6.1.1 *Localization and the neuroscience of memory*

Objection: *But look, that the hippocampus is correlated to an organism's ability to remember, especially to the working or short-term memory of the organism, is established scientific fact. That LTP is part of the neural mechanism of the remembering process is all but established scientific fact. The neuroscience of memory has been making astounding progress for decades, in humans and nonhuman animals, and has successfully established just the sort of localization of component processes you have ruled out. Doesn't this just mean that you're wrong?*

Two points are worth reiterating as a preliminary response to this worry. First, memory scientists are quick to point out that the gaps that still remain between brains and behaviors are nontrivial.[515] We have already seen that even the most optimistic reduction-oriented cognitive neuroscientists admit that any rigorous explanation of *any* cognitive capacity in terms of its neural mechanisms is still a long way off. We have also seen that memory scientists confess, over and over again, that the neural mechanisms of memory are not yet understood. Cognitive neuroscientists who work on human memory are mystified by what the mechanisms of memory could even possibly look like. Even though it is true that there are both brain areas and neural mechanisms that are widely believed to be somehow related to remembering (and with good reason, in some cases), this is far from establishing the localization of particular memory functions

[515] Tsodyks 2007, p. 71, McIntosh 2007, p. 59. Sara, writing in 2000 and quoted above in footnote 1, says "Virtually nothing is known about the physiological processes underlying the act of remembering" (Sara 2000, p. 76).

in the brain. It is clear that neuroscience has made astounding progress in the last half-century, but it is not clear that this progress has been progress toward neural localization of function.[516]

Second, remember that none of the above considerations ruled out token-token correlations, and interesting generalizations or models that might be made from these. That is, although there is no neural process that is what it is to remember, full stop, there is some neural process or collection of neural processes going on with each act of remembering. In the laboratory, very specific remembering behaviors are often correlated with very specific neural conditions. Rats injected with protein synthesis inhibitors exhibit deficits in long-term fear-conditioning.[517] This is almost certainly because protein synthesis is required for long-term potentiation and depression, which is in turn required for successful fear-conditioning.[518] This does not establish the neural mechanism of memory—it does not even establish the neural mechanism of fear-conditioning in rats. Rather, as careful memory scientists already admit, it establishes that successful LTP is a necessary condition for the normal performance of fear-conditioning in rats. That there should be such necessary conditions for constrained memory tasks is no surprise, and these in no way establish "the" residence of memory in the brain.

These being said, an enormous quantity of experimental results have demonstrated connections between specific memory phenomena and both the hippocampus (and other areas in the medial temporal lobe) and LTP or LTD (long-term depression). Some of these results concern

---

[516] Once more, see James vs. de Broca in the 1890's, and the *exact same debate* represented in Wilmes and Poeck 1993 or Plaza et al. 2009.

[517] Schafe et al. 1999.

[518] Nader et al. 2000.

LTP *in* the hippocampus,[519] and either or both of these are often taken to constitute the neural basis of remembering. Notably, the hippocampus is more likely to be baldly claimed the "seat" of memory, and LTP is more likely to be claimed to be "the mechanism" of memory, by those who do not in particular study the neuroscience of memory. All the same, these two discoveries, the hippocampus and its relation to memory phenomena and the relationship between long-term potentiation and remembering, are certainly the two biggest success stories to come out of the cognitive neuroscience of memory in the last half-century. Why, then, do cognitive neuroscientists of memory tend to continue to claim that the brain mechanisms of memory are not understood?

One reason for this is that many cognitive neuroscientists realize the difficulty in isolating a psychological function for tests of correlation, especially in human subjects. Another reason is simply that the hippocampus and its neighboring structures do not seem to constitute necessary conditions for remembering in general. A damaged hippocampus often results in various memory impairments. Some tasks that involve remembering also correlate to increased blood oxygenation levels in the human hippocampus (and the rest of the medial temporal lobe). Nonetheless, some remembering tasks are not impaired by a damaged hippocampus and some do not correlate with increased brain activity in these regions.[520] These factors have driven memory scientists to propose brain redundancies and other alternatives to the hippocampal "seat" of memory.[521] Given that the functioning of the hippocampus is not even a necessary condition for various cases of remembering, any claim to its functional localization of memory is severely diminished.

---

[519] Tim Bliss and Terje Lømo are usually credited with the discovery of LTP for their work with the hippocampus of rabbits (1973).

[520] Similar experiments have shown that right frontal cortex regions that are consistently correlated with episodic remembering fail to produce impairments when damaged (Schacter 2008, p. 67).

[521] Nadel et al. 2000.

The story is quite different for the other success story, LTP. LTP (and its counterpart LTD) are now well understood neurophysiological phenomena. The molecular mechanisms that underlie these processes are also well understood. Neurons and their connections are constituted such that simultaneous firings across synapses tend to increase (or in the case of LTD, decrease) sensitivity to future stimuli at that synapse. Even more promisingly, the functional role of LTP and LTD in the neural network of the brain is very like the functional role played by the strengthening and weakening of connections in an artificial neural network. This strengthening and weakening is at the heart of what allows artificial neural networks to learn to distinguish or replicate complex patterns and stimuli. Furthermore, we know that successful LTP is a necessary condition for at least some memory tasks in organisms.

This might almost sound like we have found the neural mechanism of memory (and it is often treated as such, though again not as often anymore by researchers who focus on the neuroscience of memory) until we realize that successful LTP is very likely to be a necessary condition for *almost every cognitive endeavor*. In their article *LTP and LTD: an embarrassment of riches*, Robert Malenka and Mark Bear cite this ubiquity as the reason that these neural processes can no longer be taken as mechanisms of memory.

> We would also argue that it is no longer particularly productive to debate the generic question of whether LTP and LTD are cellular/synaptic mechanisms for memory. LTP and LTD are experimental phenomena, which can be used to demonstrate the repertoire of long-lasting modifications of which individual synapses are capable. It is a daunting task to demonstrate that identical synaptic modifications due to the same mechanisms underlying some form of LTP or LTD occur in vivo in response to experience. It is even more difficult to prove that these LTP or LTD-like modifications subserve essential functional roles.[522]

Long-term potentiation and depression are simply qualities of neural connections. They are interesting qualities, and we should (and will) continue to study their operation and their effect on

[522] Malenka and Bear 2004, p. 5.

a brain and organism, but they cannot serve as "the mechanism" of memory, unless, as James suggested about language, we take the description of remembering to be the description of the whole human subject in miniature.

The neurobiology of remembering should, and will, continue to be studied. To conceive of the hidden neural process that implements remembering as analogous to the hidden physiological process that implements, for example, digestion, is to misconstrue what memory is. We may, and probably will, find many more instances of interesting correlations between particular remembering contexts and certain neurophysiological processes. However, remembering is not the sort of process that is *implemented by* a particular internal, physiological mechanism.

6.1.2 *Metaphor*

Objection: *Alright, so the event or item isn't actually stored in my mind or brain the way that my canned goods are stored in my pantry, but I would have conceded this if pressed anyway. My mind isn't very much like a pantry and past events, or ideas about them, aren't very much like canned goods. Isn't storage just a metaphor we use to describe the way that we can revive past events or experiences? In philosophy in particular, there must be many cases where this metaphor is employed benignly. Are you really trying to persuade philosophers which metaphors they can and can't use?*

It is true that the storage metaphor did not begin with the computer metaphor. Storage and retrieval is a natural way to casually describe the operations of memory in some instances, and one need not be influenced by computationalism or attempting to model human memory in order to be inclined to employ such a description. In addition, although this project has focused on *disanalogies* between stored items and memories, there are clearly analogies between these as well. In some cases of remembering, looking for and finding a slightly degraded memory of a long-past event really is a lot like looking for and finding a slightly degraded item that was put into storage long ago. If these examples of remembering are the examples we focus on, the storage metaphor can seem not only benign but particularly apt.

Outside of memory science or the philosophy of memory, there must have been countless occasions on which such metaphors were employed without any disastrous conceptual effects. Even within the science and philosophy of memory, storage and retrieval models have probably played harmless roles in many cases: experiments that aim to determine, for example, the rates at which people forget learned information or the effects of certain drugs on memory impairment in animal (or human) subjects might often deal very little in the models or theoretical structure of remembering. While there may not be such a thing as *pure*, theory-free data, some experimentation is certainly more theory-free than other experimentation. In experiments where the memory store was not playing a special role *qua* memory store, the damage has very probably been quite limited.

All of that being said, the short answer is still 'yes'. That is, I am trying to persuade philosophers to be very cautious about the metaphors they use in characterizing human memory, and I am trying to destabilize the tendency to automatically conceive of memory in terms of storage and retrieval. Our choice of metaphor very often shapes the way that we think of the phenomenon in question, and the human ability to remember is closely related to many areas of difficult and contentious philosophical interest and research. Given that those who study memory, even those who do not offer revisions quite as dramatic as those above, explicitly reject the fixed, stored memory trace as an adequate model of human memory, philosophers and others should not disregard this warning when we invoke memory in other discussions.[523]

Additionally, the historical considerations in chapter two were supposed to show that these computationalist assumptions can lie undetected in the language of storage and retrieval, even

---

[523] Bennett and Hacker worry that the storage metaphor in particular causes confusion by "being taken to be what it is merely a metaphor for" (Bennett and Hacker 2003, p. 159).

when this language comes from those who would otherwise seek to reject such claims. A theoretical commitment to storage entails numerous difficulties. I have tried to show that the language we couch characterizations of memory in really does affect our theoretical commitments, and sometimes in unseen and significant ways. Perhaps more than anyone else, philosophers should be concerned about the ways that theoretical commitments can be smuggled into language; seemingly benign metaphors can be instruments of just this.

6.1.3 *The reality and nature of inner processes*

Objection: *You talk a lot about the "remembering behavior" and the many "expressions" that remembering can take, and you generally impugn inner processes. Behaviorism is dead, right? I can, right here and now, recall the face of an old friend. I can bring this image to mind without any outward expressions or behavior whatsoever, and it seems for all the world like a hidden, inner process of just the variety you've ruled out. If remembering isn't an inner process, what am I doing?*

Hopefully some of the considerations above concerning hidden processes, especially in sections 2.3.3, 3.3.1, 4.1.3, and 4.4.4 helped to clear up some of the confusions that could arise here. I have not denied our ability to form mental images of remembered scenes, in fact this ability was explicitly endorsed by both Ayer and Wittgenstein *as* they were arguing against memory as an inner process. We certainly can, and often do, remember by visualizing something that we encountered in the past. There are two immediate reminders needed concerning this kind of "inner process". First, this is not a non-experienced, inner cognitive-level process. It is experienced. Mental imagery is a phenomenal process, and an interesting one. Whatever most memory scientists mean by the processes of remembering—such as encoding or retrieval—these are not taken to be phenomenal processes. Indeed, I take it as obvious and uncontroversial that sometimes, in the course of remembering, I also experience inner, mental processes of imagery. The remembrance that shares a second-order resemblance with the past event (though not with any interim substrate) can certainly be such an inner image. Even without any further analysis, it seems relatively uncontroversial to claim that this kind of inner remembrance bears isomorphism to the past event that is recalled.

Second, there was a claim about this kind of imagery, but the claim was simply that this cannot be "the" process of remembering, that this kind of imagery is not necessary for remembering. This lack of necessity of an inner image is easy to demonstrate, it has been argued. We might sometimes remember with an inner image, but we often do not. The image has no special place in remembering. We might remember an image, and then describe it. But we may also describe the past, and then imagine it. Neither of these ways of remembering enjoys priority over the other.

Because Wittgenstein is sometimes characterizes as a sort of naïve behaviorist, it is important to get clear about the arguments on the table. It seems untenable to ascribe naïve behaviorism to Wittgenstein: he discusses the phenomenology and role of inner images often and in various contexts. He certainly admits the reality of such phenomena. *This* kind of inner mental process has not been denied. It is true that in undermining the notion that phenomenal imagery is essential to processes like remembering, Wittgenstein is, in a way, de-emphasizing inner processes like imagery or "inner speech", but this de-emphasis does not equate with denial. Sometimes the existence of inner imagery is taken as warrant for the whole of the hidden cognitive level, conscious and unconscious, and as warrant for cognitive-level explanations of particular phenomena. This topic is too vast and complex to justly treat in this project, but I might offer a sketched response.

The inference escapes me. There are inner, conscious, experienced mental phenomena. I can undoubtedly picture the face of an old friend, or mentally hum a tune, or sit here and work out relatively complex arithmetic without any overt indications of my thoughts. In this conscious sense, inner mental phenomena certainly exist. The inference to the existence software-level hidden and unconscious processes seems unwarranted. One might say that these are just like inner mental processes, only not conscious, but I do not find this illuminating. It is also true that there are ways to deliver stimuli such that the subject does not consciously report their perception, and yet the stimulus affects the subject's future reactions in various and predictable ways. This might be described as a kind of "subconscious" perception, but there is no need or room for cognitive-level remembering processes in the account of these instances.

Although we have seen many arguments and results that are generally hostile toward unconscious, inner, cognitive processes, none of these were supposed to decisively demonstrate anything about these kinds of processes outside of memory. The arguments above show that inner subprocesses of remembering are extraneous, unhelpful, conceptually problematic, and fit the data poorly. In the science and philosophy of remembering, is has been argued, inner cognitive processes have been overrated and overemphasized, and they have not been critically investigated enough. These investigations undermine the inner subprocesses of memory, but they do not undermine the existence of phenomenal mental processes of memory, nor of inner subprocesses in other areas of cognition.

6.2 Implications

Many of the implications of these revisions to our conception of memory for the sciences of memory have already been noted. The following is a summary of some of the more important takeaways from these revisions. There are important implications both inside and outside of the cognitive sciences of memory. Cognitive scientists of memory may worry that the picture of remembering we are left with, which includes neither a neural trace nor component subprocesses, is one that simply cannot be studied empirically. Nothing could be further from the truth. The picture of memory presented here is not a reductionist picture: there is no neural process that is remembering, or that implements the process that is what it is to remember. This anti-reductionism (and again, it is a local anti-reductionism, concerning only the cognitive capacities of remembering) can only be taken as anti-scientific on a ruthlessly reductionist view of the scientific enterprise. An overly reductionist conception of the target of some inquiry can just as easily obscure the way forward as can an overly non-reductionist conception. Hilary Putnam writes:

> Not only does rejecting reductionist pictures not entail abandoning serious scientific research but, in fact, it is those pictures that often lead researchers to misconceive the empirical problems.[524]

Just so. In the science of memory, the automatic assumption that whatever processes there were to study must ultimately be, or be implemented by, neurophysiological processes (that are in turn identical with or implemented by biochemical processes) is just one of the assumptions that has contributed to these many impasses and difficulties.

We need not work out the relationship between traces, subprocesses, and scientific progress *a priori*, however. The evidence from the last fifty years of memory science directly contravenes the supposition that scientific progress in memory studies is constituted by revealing traces and component subprocesses. We have enjoyed fifty years of successful memory science, in many, many ways, and we still have no trace or non-controversial subprocesses to show for it. That is, none of these considerations provide reasons to reject the empirical work that has been done on the ways people actually remember, the ways remembering actually works, or what is actually happening in the brain in various circumstances. On the contrary, the revelations from these studies have provided direct evidence for the theoretical revisions suggested here. Memory

---

524 Putnam 1999, p. 174.

science has generated an enormous body of well-researched results; it is the theory and interpretation of these results that deserve scrutiny and revision.[525]

Within the science of memory, there are thus several constructive implications that stem from this revised conceptual framework for memory. Perhaps the most striking of these is the revisions to strategies for connecting the neural and the behavioral or phenomenal levels of study. Cognitive experience results in brain changes. Remembering behavior and phenomena depend on neural conditions. Some particular neural conditions and processes are correlated to particular cognitive tasks of remembering. Even in simplified cases, the human behavior of remembering is complex. Nonetheless, some of these complexities can be reliably modeled and predicted. Our capacities to remember the past persist under certain conditions. Under other conditions, some of these capacities fail to persist. The conditions—neurobiological, psychological, or otherwise—that allow for persistence of particular instances and varieties of remembering are also complex, but these also lend themselves to modeling and prediction. Connecting any or all of these levels and phenomena is no easy task, but nor is it impossible. Jonathan Trigg and Michael Kalish tell the following story about action and its enabling conditions in order to demonstrate a Wittgensteinian approach to the connection between neural conditions and cognition:

> Now consider Paul's opening a door. For the purpose of argument, suppose that this action consists in his reaching out, taking hold of the doorknob, holding onto the doorknob, turning the doorknob, and pulling the door toward him. There is reason to say that there are not five actions here but one—Paul's opening the door—but nevertheless, in opening the door, he does all these things, or he opens the door by doing all these things. When he opens the door, all sorts of events happen in his arm and hand including various muscle contractions; but contracting his muscles (etc.) is not something Paul does—he does not, for example, turn the doorknob by contracting various muscles in his hand and wrist. Here we could say that his reaching out, his taking hold of the doorknob, etc. are partially constitutive of his action of opening the door; and that the muscle contractions in his arm

[525] This is not to suggest that there is any kind of complete disconnect between theory and data. Indeed, although we have learned many things from, for example, experiments that were designed to test whether impairments were due to encoding failure or retrieval failure (including the lessons above, which have been learned by the study of just these experiments), *we have not learned whether impairments were due to encoding failure or retrieval failure*.

> are causally necessary enabling conditions of this action. Note, again, that these muscle contractions do not cause the action Paul performs (nor any of the things he does in performing this action), but if they did not happen, Paul would be quite unable to open the door, at least by doing the things we have described him as doing in this case.[526]

Trigg and Kalish are walking a fine line here in characterizing the muscle contractions as "causally necessary" but not "causal", but their point is well taken.[527] Although actions are notoriously tricky to individuate, Paul did not contract his muscles so that he could open the door, and it was not the contraction of his muscles that committed the action of opening the door. Rather, Paul opened the door, and various muscular contractions were necessary for him to succeed. Later, Paul may remember opening the door, and various neurobiological processes and conditions may be necessary for him to succeed in remembering this. The neurobiological conditions that enable various remembrances are already under investigation, and have been so for decades. We need not reduce components of memory to their neural counterparts—no more than we need (or can) reduce components of chess strategies to their neural (or physical) counterparts.

We have seen above that models of memory that adequately reflect the phenomena will be those models, like Michaelian's generative memory, that make the dynamic and constructive nature of human remembering an essential component. We have also seen that successful models will be those models, as Dudai's begins to do with 'persistence' models, that jettison the language and conceptual trappings of storage and retrieval. Memory scientists already recognize, in name, that archival models have been undermined by empirical and philosophical considerations; now this insight need only be expressed with conviction. We have also seen that successful memory

---

526 Trigg and Kalish 2011, p. 408.

527 This is presumably in response to the fact that Paul's beating heart is also a physically necessary condition for his opening the door, but Trigg and Kalish want to fairly characterize the special relationship the muscular contraction has with the door-opening, over and above the sort of necessary condition the heart provides. There may be no particular problem with using 'causally necessary' to denote this relationship, but the distinction between merely necessary and *causally* necessary will be less sharp than that between causally necessary and *causal*.

models will pay attention not just to the resources from connectionism and dynamical systems, but also to the way that these are wielded by researchers who primarily work on neural networks and models of dynamical systems.

Another important implication for memory science is the lesson concerning treating models *as models*. That is, memory scientists already sometimes talk as though elements of their theories are merely explanatory models. We can name the rate at which the capacity to remember something diminishes 'the forgetting curve' without causing much confusion—nobody wonders where the forgetting curve is found in the brain, or how forgetting curves are neurally implemented. The forgetting curve is a purely theoretical construct. It can be used in explanations, and it can be refined to make better and better predictions, but it is an element in a model. Memory scientists can also posit, for example, the "visual buffer", if by this they only mean to model the human capacity to remember visual information for a little while. This framework only gets into conceptual difficulty when the visual buffer is reified, and cognitive scientists wonder which part of the brain the visual buffer can be found in. This, as we have seen, is no more sensible than wondering what materials constitute the earth's axis. The cognitive psychologist's "visual buffer" originated as a theoretical element in a model.[528] Forty years and several conceptual confusions later, cognitive neuroscientists have identified the "visual buffer" as a brain area.[529] Reliably, cognitive scientists are now puzzled as to why data from brain imaging and remembering tasks "offer a challenge to the assumed overlap between perception and visual imagery and to the assumed role of the visual buffer as an input gateway between perception and the cognitive

[528] Baddeley and Hitch 1974.

[529] Stephen Kosslyn writes, "I group the topographically organized visual areas of the cortex into a single functional structure, which I call the visual buffer" (Kosslyn 2005, p. 336).

system."[530] This challenge is on offer because what was a theoretical construct expressing the many ways in which our ability to recall visual information quickly fades has become a mechanism in the brain. This is exactly the sort of confusion that can be ameliorated by remembering to treat models as models.

There are also implications outside of memory science. In other areas of cognitive science, the memory store is still sometimes treated as a given. As we have already seen, Kosslyn's model of perception and imagery depends on a memory store in just the ways that were criticized in chapter four. If memory does not involve storage, these models of perception will have to be dramatically revised. Arguments again subsumption architectures and other nonrepresentational approaches to cognition level what is sometimes called the "scope objection"[531]—that such nonrepresentational systems will be unable to "scale up" to cognitive tasks like memory storage. If memory is not storage, then there is one more large and important class of cognitive capacities that needn't trouble the proponent of nonrepresentational cognition.

There are also implications for various points of discussion in philosophy. As we have already seen, accounts of attention and consciousness such as Prinz's depend on the operations of working memory. Prinz makes explicit use of storage and retrieval in this part of his account, relying on what is stored—over and above what is experienced—to determine what we can be conscious of.[532] Theories of content such as Burge's make explicit use of the memory store, relying on memory's ability to reproduce preserved mental states in order to constitute

---

530 Logie and Van der Meulen 2009, p. 11.

531 Hutto and Myin 2013, p. 38.

532 Prinz 2012, p. 101.

representational states.[533] If memory does not involve storage, neither of these accounts will survive unrevised. Memory, and the specific way memory functions is also important to a collection of epistemological problems.[534] In most cases, the "semantic memory" of memory researchers is just the knowledge-that of epistemologists. Kourken Michaelian provides an argument against a "preservationist" view about the epistemology of memory—that memory can preserve but not generate justification—based on the generative ways that memory actually works.[535]

Lastly, this revised account of remembering may have implications outside of theoretical research altogether. What goes on when we remember is intimately tied to how we learn and act as human beings. Memory science often has implications for education. Decades ago, memory scientists were codifying the various ways that multimodal processing increased retention in memory tasks.[536] Years later, educators and educational researchers are applying these ideas in the classroom.[537] If remembering does not involve storing, and the processes of memory are more like the processes of shaping clay than they are like the processes of saving work on a computer, there may be ways to implement these insights in education and learning. At least in some

---

533 Burge 2010, p. 63.

534 Thomas Senor writes, "Virtually all of what we know (or are justified in believing) at any given time resides in memory. However, epistemology has been focussed almost exclusively, even if implicitly, on the epistemology of belief formation — of *coming* to believe a proposition" (Senor 2014).

535 Michaelian 2011a, §3. Although he is does not attend to the empirical details of memory as closely, McGrath (2007) makes a similar argument against both preservationism and evidentialism based on the way that memory works.

536 See Kapur et al. 1994 for an influential treatment of different "levels of processing".

537 For a recent and explicit example of this influence, see Moreno and Mayer 2007.

contexts, a non-archival conception of memory will imply a preference for learning techniques that are more like training than they are like uploading.

6.3 Further investigations

The foregoing is not a model of memory, but rather a framework within which memory models might be constructed such that they avoid the empirical, conceptual, and philosophical problems that memory science is presently encountering. The work remains to build models of memory and of remembering phenomena that fit into this framework. This is largely the work of cognitive psychologists and neuroscientists, of course: modeling memory phenomena requires the acute and scientific study of memory phenomena. Modeling memory phenomena well, however, also requires careful attention to the conceptual groundwork the model is based on. As we have already seen, many contemporary accounts of memory have developed in response to the philosophical and empirical difficulties from chapters three and four. As we have also seen, many of these accounts have responded insufficiently, maintaining the language and basis of storage and retrieval. Much of the work that remains consists in revisiting these accounts and extracting computationalist assumptions from them—much as the previous chapter did by extracting traces from Michaelian's account of generative memory. Given that memory scientists have been, in recent years, clamoring for conceptual investigations and grasping about for alternative metaphors and frameworks, memory science in the 21st century provides fertile ground for philosophical work. Further collaboration between philosophers and scientists, as well as further integration across the many disciplines that make up "memory studies", will continue to build a common ground that could prove to be one of the most fruitful in the inherently multidisciplinary endeavor that is cognitive science.

This Wittgensteinian study of the cognitive science of memory also provides a launch point for further investigation of Wittgenstein's philosophy as it is applied to contemporary studies of mind and cognitive science. John Sutton's recent article on Wittgenstein and memory expresses the worry that Wittgenstein scholars are sometimes guilty of operating at too far a remove from the actual workings of cognitive science.[538] Sutton worries about this especially because he does believe that Wittgenstein's insights have something to offer to memory science, and to cognitive science in general. If philosophers influenced by Wittgenstein believe that his work has something to offer to our understanding of mentality and cognition, the onus is on these philosophers to take up Wittgensteinian investigations of the sciences and philosophy of cognition. Furthermore, as surveyed above, although Wittgenstein is counted among the most important philosophers of mind and language in the twentieth century, noncontroversial characterizations of his contributions to philosophy of mind remain elusive. None of the above is supposed to contain or express a systematic "Wittgensteinian vision of mind", but it is studies like these that will sharpen our understanding of what Wittgenstein has to offer to contemporary studies of mind and cognition. There are probably many other cognitive capacities for which it would be fruitful to consider accepting psychological indeterminacy, or for which it may aid our conceptual approach to ask which of the purported mental states in question had durations. Not only can these help to clarify areas of cognitive science, but they can also add up to a Wittgensteinian take on the mind. Understanding exactly what these individual studies amount to, and how they comprise a Wittgensteinian approach to cognition, remains an interesting, further task.

Lastly, nonrepresentational approaches to cognition have been flourishing in recent years, but none of these have explicitly focused on memory in a sustained fashion. The encoded memory

---

[538] Sutton 2014.

taken for granted by standard models is an obstacle for nonrepresentational or minimally representational theories of cognition. Ramsey writes that "according to many, any functional architecture that is causally responsible for the system's performance can be characterized as *encoding* the system's knowledge base."[539] Ramsey goes on to express dismay that this means that a nonrepresentational account of mental capacities "is not simply implausible—it is virtually *inconceivable.*"[540] Although representation *per se* played little role in the conceptual revisions above, there are no stored representations in a model of memory that precludes storage. Some proponents of nonrepresentational cognition may be dogmatically against mental representation, but many researchers in cognitive science have an attitude toward representation that is reminiscent of Morgan's Canon: do not invoke representational mechanisms when nonrepresentational mechanisms will do the job. Modern embodied and enactivist approaches have focused on deflating the need for representation in perception and problem-solving, but have not focused on remembering. Accounts of memory that do not invoke stored representations may be instructive. Whatever their allegiances concerning fundamental representation, almost all researchers in cognitive science are interested to see how far this line can be pushed. That human remembering may proceed without storing anything—certainly without storing representations—demonstrates one more large and important class of cognitive capacities that does not require inner mental representations. If this is so, then the traditional, computationalist role of the inner mental item in cognition has been significantly diminished.

Human rememberers can get along without storing or retrieving anything. Memory scientists who have run aground on latent computationalist assumptions have been, for years,

---

539 Ramsey 2007, p. 2, emphasis added.

540 *Ibid*., original emphasis.

building an assortment of cases against archival models and metaphors for years. Wittgenstein and other philosophers have been, for decades, presenting arguments against the efficacy of, and demand for, the storage and retrieval of an inner mental item in order to explain remembering. A collection of research programs in and outside of the study of memory have been successfully moving away from computationalist models, and making real progress toward understanding human memory outside of computationalist terms. The lingering emphasis on encoding, storage, and retrieval may only serve as an impediment. Remembering can and should be studied, modeled, and explained without recourse to the trappings of the computer metaphor.

BIBLIOGRAPHY

Anscombe, G. E. M. (1957). *Intention*. Harvard University Press.

Austin, J. L. (1975). *How to do things with words*. Oxford university press.

Ayer, A. J. (1956). *The Problem of Knowledge*. Penguin Group (USA) Incorporated.

Babiloni, F., & Astolfi, L. (2012). Social neuroscience and hyperscanning techniques: Past, present and future. *Neuroscience & Biobehavioral Reviews*.

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556–559.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*(10), 829–839.

Baddeley, A. D. (2007). Working memory: Multiple models, multiple mechanisms. In H. L. Roediger III, Y. Dudai, & S. M. Fitzpatrick (Eds.), *Science of Memory: Concepts* (pp. 151–153). Oxford University Press.

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation*, *8*, 47–89.

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.*, *59*, 617–645.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge University Press.

Bartsch, T. (2012). *The Clinical Neurobiology of the Hippocampus: An integrative view*. OUP Oxford.

Bechtel, W. (2009). Explanation: mechanism, modularity, and situated cognition. *The Cambridge Handbook of Situated Cognition*, 155–170.

Bekinschtein, P., Cammarota, M., Igaz, L. M., Bevilaqua, L. R., Izquierdo, I., & Medina, J. H. (2007). Persistence of long-term memory storage requires a late protein synthesis-and BDNF-dependent phase in the hippocampus. *Neuron*, *53*(2), 261–277.

Bennett, M. R., & Hacker, P. M. S. (2003). *Philosophical Foundations of Neuroscience*. Blackwell Oxford.

Bennett, R., Dennett, D., Hacker, P., & Searle, J. (2013). *Neuroscience and Philosophy: Brain, Mind, and Language*. Columbia University Press.

Bergson, H. (1896). *Matter and Memory*. Courier Dover Publications.

Bermúdez, J. L. (2003). *Thinking without Words*. Oxford University Press.

Bernecker, S. (2001). Russell on mnemic causation. *Principia*, *5*, 149–185.

Bernecker, S. (2008). *The Metaphysics of Memory*. Springer.

Bernecker, S. (2009). *Memory: A philosophical study*. OUP Oxford.

Bickhard, M. H., & Terveen, L. (1996). *Foundational issues in artificial intelligence and cognitive science: Impasse and solution*. Elsevier.

Bickle, J., Mandik, P., & Landreth, A. (2012). Philosophy of neuroscience. In (E. N. Zalta, Ed.) *The Stanford Encyclopedia of Philosophy*. Retrieved from http://plato.stanford.edu/archives/sum2012/entries/neuroscience/

Bischof-Köhler, D. (1985). Zur phylogenese menschticher motivation [On the phylogeny of human motivation]. *Emotion Und Reflexivitat. Urban & Schwarzenberg*, 3–47.

Bliss, T. V., & Lømo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *The Journal of Physiology*, *232*(2), 331–356.

Bloch, M. (1998). *How we think they think: Anthropological approaches to cognition, memory, and literacy*. Westview Press Oxford.

Block, N. (1990). Can the mind change the world? In G. Boolos (Ed.), *Meaning and method: Essays in honor of Hilary Putnam* (pp. 137–170). Cambridge University Press.

Boncompagni, A. (2013). Enactivism and the “explanatory trap”. A Wittgensteinian perspective. *Methode-Analytic Perspectives*, *2*(2), 27–49.

Brett, M., Johnsrude, I. S., & Owen, A. M. (2002). The problem of functional localization in the human brain. *Nature Reviews Neuroscience*, *3*(3), 243–249.

Bringsjord, S. (1998). Computationalism is dead; now what? *Journal of Experimental & Theoretical Artificial Intelligence*, *10*(4), 393–402.

Broad, C. D. (1925). *The Mind and its Place in Nature*. Routledge & Kegan Paul London.

Brockmeier, J. (2002). Remembering and forgetting: Narrative as cultural memory. *Culture & Psychology*, *8*(1), 15–43.

Brockmeier, J. (2010). After the archive: Remapping memory. *Culture & Psychology*, *16*(1), 5–35.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, *47*(1), 139–159.

Brown, J. H. (1968). An uncommon type of transient loss of memory. *Canadian Medical Association Journal*, *98*(18), 878.

Buckner, R. L. (2007). Memory systems: An incentive, not an endpoint. In H. L. Roediger III, Y. Dudai, & S. M. Fitzpatrick (Eds.), *Science of memory: Concepts* (pp. 359–364). Oxford University Press.

Buckner, R. L., & Wheeler, M. E. (2001). The cognitive neuroscience of remembering. *Nature Reviews Neuroscience*, *2*(9), 624–634.

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

Burton, J. (2008). Bergson's non-archival theory of memory. *Memory Studies*, *1*(3), 321–339.

Bussey, T. J. (2004). Multiple memory systems: Fact or fiction? *The Quarterly Journal of Experimental Psychology B: Comparative and Physiological Psychology, 57B*(1), 89-94.

Cabeza, R., & Moscovitch, M. (2013). Memory Systems, Processing Modes, and Components Functional Neuroimaging Evidence. *Perspectives on Psychological Science*, *8*(1), 49–55.

Carretero, M., & Solcoff, K. (2012). Commentary on Brockmeier's remapping memory: The relation between past, present and future as a metaphor of memory. *Culture & Psychology*, *18*(1), 14–22.

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

Chalmers, D. J. (1996). *The Conscious Mind: In search of a fundamental theory*. Oxford University Press.

Chemero, A. (2011). *Radical embodied cognitive science*. MIT press.

Churchland, P. M. (1993). State-space semantics and meaning holism. *Philosophy and Phenomenological Research*, 667–672.

Clark, A. (1998). *Being there: Putting brain, body, and world together again*. MIT press.

Clark, A. (2001). *Mindware: An Introduction to the Philosophy of Cognitive Science*. Oxford University Press.

Clark, A. (2010). Memento's revenge: The extended mind, extended. In R. Menary (Ed.), *The Extended Mind* (pp. 43–66). Cambridge, MA: MIT Press.

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

Clayton, N. S., & Dickinson, A. (1998). Episodic-like memory during cache recovery by scrub jays. *Nature*, *395*(6699), 272–274.

Cockburn, D. (2001). Memories, traces and the significance of the past. *Time and Memory: Issues in Philosophy and Psychology*.

Cohen, G., & Conway, M. A. (2007). *Memory in the real world*. Psychology Press.

Copeland, B. J. (1996). What is computation? *Synthese*, *108*(3), 335–359.

Copeland, B. J., & Proudfoot, D. (2010). Deviant encodings and Turing's analysis of computability. *Studies in History and Philosophy of Science Part A*, *41*(3), 247–252.

Costall, A. (2013). The Unconscious Theory in Modern Cognitivism. *A Wittgensteinian Perspective on the Use of Conceptual Analysis in Psychology*, 312.

Craik, F. I., & Jennings, J. M. (1992). Human memory. In T. Salthouse (Ed.), *The handbook of aging and cognition* (pp. 51–110). Lawrence Erlbaum Associates.

Craik, F. I. M. (2007). Encoding: A cognitive perspective. In H. L. Roediger III, Y. Dudai, & S. M. Fitzpatrick (Eds.), *Science of memory: Concepts* (pp. 129–135). Oxford University Press.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684.

Cummins, R. (1991). The role of representation in connectionist explanations of cognitive capacities. *Philosophy and Connectionist Theory*, 91–114.

Cummins, R. (2000). "How does it work?" versus "what are the laws?": Two conceptions of psychological explanation. *Explanation and Cognition*, 117–144.

D'Esposito, M. (2007). From cognitive to neural models of working memory. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 761–772.

D'Esposito, M., Detre, J. A., Alsop, D. C., Shin, R. K., Atlas, S., & Grossman, M. (1995). The neural basis of the central executive system of working memory. *Nature*, *378*(6554), 279–281.

Dalla Barba, G. (2001). Beyond the memory-trace paradox and the fallacy of the homunculus. A hypothesis concerning the relationship between memory, consciousness and temporality. *Journal of Consciousness Studies*, *8*(3), 51–78.

Danziger, K. (2008). *Marking the Mind: A History of Memory*. Cambridge University Press.

Davidson, D. (1975). Thought and talk. *Mind and Language*, *1975*, 7–23.

Davidson, D. (1982). Rational animals. *Dialectica*, *36*(4), 317–327.

Démonet, J.-F., Thierry, G., & Cardebat, D. (2005). Renewal of the neurophysiology of language: functional neuroimaging. *Physiological Reviews*, *85*(1), 49–96.

Dennett, D. C. (1981). Three kinds of intentional psychology. In R. Healey (Ed.), *Reduction, Time and Reality* (pp. 163–186). New York, NY: Cambridge.

Dennett, D. C. (1991). Real patterns. *Journal of Philosophy*, *88*(1), 27–51.

Dennett, D. C. (2006). The Frame Problem of AI. *Philosophy of Psychology: Contemporary Readings*, 433.

Descartes, R. (1991). *The Philosophical Writings of Descartes: Volume 3, The Correspondence*. (J. Cottingham, R. Stoothoff, & D. Murdoch, Trans.). Cambridge University Press.

Deutscher, M. (1989). Remembering "remembering." In *Cause, Mind, and Reality* (pp. 53–72). Springer.

Dokic, J. (2001). Is memory purely preservative? *Time and Memory. Issues in Philosophy and Psychology.*, 213–232.

Draaisma, D. (2000). *Metaphors of memory: A history of ideas about the mind*. Cambridge University Press.

Dreyfus, H. L. (1972). *What computers can't do: A critique of artificial reason*. Harper & Row New York.

Dreyfus, H. L. (1981). From micro-worlds to knowledge representation: AI at an impasse. *Mind Design*, 161–204.

Dreyfus, H. L. (2007). Why Heideggerian AI failed and how fixing it would require making it more Heideggerian. *Philosophical Psychology*, *20*(2), 247–268.

Dreyfus, S. E., & Dreyfus, H. L. (1980). *A five-stage model of the mental activities involved in directed skill acquisition*. DTIC Document.

Dudai, Y. (2002). Molecular bases of long-term memories: a question of persistence. *Current Opinion in Neurobiology*, *12*(2), 211–216.

Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram? *Annu. Rev. Psychol.*, *55*, 51–86.

Dudai, Y. (2006). Reconsolidation: the advantage of being refocused. *Current Opinion in Neurobiology*, *16*(2), 174–178.

Dudai, Y., Roediger III, H. L., & Tulving, E. (2007). *Science of Memory: Concepts*. Oxford University Press.

Eichenbaum, H. (1997). Declarative memory: Insights from cognitive neurobiology. *Annual Review of Psychology*, *48*(1), 547–572.

Eichenbaum, H. (2007). Persistence: Necessary but not Sufficient. In H. L. Roediger III, Y. Dudai, & S. M. Fitzpatrick (Eds.), *Science of memory: concepts* (pp. 193–198). Oxford University Press.

Elman, J. L. (1991). Representation and structure in connectionist models. In *Cognitive Models of Speech Processing*. MIT Press.

Feeney, M. C., Roberts, W. A., & Sherry, D. F. (2009). Memory for what, where, and when in the black-capped chickadee (Poecile atricapillus). *Animal Cognition*, *12*(6), 767–777.

Figdor, C. (2009). Semantic Externalism and the Mechanics of Thought. *Minds and Machines*, *19*(1), 1–24.

Fillenbaum, S. (1966). Memory for gist: Some relevant variables. *Language and Speech*, *9*(4), 217–227.

Fodor, J. A. (1975). *The Language of Thought*. Harvard University Press.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*(1), 3–71.

Foster, J. K., & Jelicic, M. E. (1999). *Memory: Systems, process, or function?* Oxford University Press.

Frongia, G. (1995). Wittgenstein and Memory. In *Wittgenstein: Mind and Language* (pp. 263–277). Springer.

Gabrieli, J. D. (1998). Cognitive neuroscience of human memory. *Annual Review of Psychology*, *49*(1), 87–115.

Gabrieli, J. D., Poldrack, R. A., & Desmond, J. E. (1998). The role of left prefrontal cortex in language and memory. *Proceedings of the National Academy of Sciences*, *95*(3), 906–913.

Gallistel, C. R., & King, A. P. (2011). *Memory and the computational brain: Why cognitive science will transform neuroscience*. John Wiley & Sons.

Gardiner, F. M., Craik, F. I., & Bleasdale, F. A. (1973). Retrieval difficulty and subsequent recall. *Memory & Cognition*, *1*(3), 213–216.

Garson, J. (2012). Connectionism. In (E. N. Zalta, Ed.) *The Stanford Encyclopedia of Philosophy*. Retrieved from http://plato.stanford.edu/archives/win2012/entries/connectionism/

Gathercole, S. E. (2007). Working memory: what it is, and what it is not. In H. L. Roediger III, Y. Dudai, & S. M. Fitzpatrick (Eds.), *Science of memory: concepts* (pp. 155–158). Oxford University Press.

Giunti, M. (1996). Beyond computationalism. In *Proceedings of the 18th annual conference of the Cognitive Science Society* (pp. 71–75). L. Erlbaum Associates, Mahwah, NJ.

Glock, H.-J. (2013). Animal minds: Philosophical and scientific aspects. *A Wittgensteinian Perspective on the Use of Conceptual Analysis in Psychology*, 130.

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, *66*(3), 325–331.

Gold, P. E., & King, R. A. (1974). Retrograde amnesia: storage failure versus retrieval failure. *Psychological Review 81*(5), 465-469.

Goldman, A. I. (1978). Epistemology and the Psychology of Belief. *The Monist*, *61*(4), 525–535.

Griffiths, D., Dickinson, A., & Clayton, N. (1999). Episodic memory: what can animals remember about their past? *Trends in Cognitive Sciences*, *3*(2), 74–80.

Gross, C. G. (2002). Genealogy of the "grandmother cell." *The Neuroscientist*, *8*(5), 512–518.

Gryz, J. (2013). The Frame Problem in Artificial Intelligence and Philosophy. *Filozofia Nauki*, *21*(2 (82)), 15–30.

Hacker, P. M. (2013). Prologue: Wittgenstein's Philosophy of Psychology as a Critical Instrument for the Psychological Sciences. In T. Racine & K. L. Slaney (Eds.), *A Wittgensteinian Perspective on the Use of Conceptual Analysis in Psychology* (p. 10). Palgrave Macmillan.

Hamilton, A. (2009). Memory and self-consciousness: Immunity to error through misidentification. *Synthese*, *171*(3), 409–417.

Hampton, R. R. (2001). Rhesus monkeys know when they remember. *Proceedings of the National Academy of Sciences*, *98*(9), 5359–5362.

Hardt, O., Wang, S.-H., & Nader, K. (2009). Storage or retrieval deficit: the yin and yang of amnesia. *Learning & Memory*, *16*(4), 224–230.

Harré, R. (2005). *Wittgenstein and psychology: A practical guide*. Ashgate Publishing, Ltd.

Hasselmo, M. E. (2007). Encoding: Models linking neural mechanism to behavior. In H. L. Roediger III, Y. Dudai, & S. M. Fitzpatrick (Eds.), *Science of Memory: Concepts* (pp. 123–128).

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. Psychology Press.

Heil, J. (1978). Traces of things past. *Philosophy of Science 45*(1), 60-72.

Henke, K. (2010). A model for memory systems based on processing modes rather than consciousness. *Nature Reviews Neuroscience*, *11*(7), 523–532.

Hoerl, C., & McCormack, T. (2001). Perspectives on Time and Memory An Introduction. *Time and Memory: Issues in Philosophy and Psychology*, 1–33.

Horst, S. (2011). The Computational Theory of Mind. In *Stanford Encyclopedia of Philosophy*. Retrieved from http://plato.stanford.edu/archives/spr2011/entries/computational-mind/

Hume, D., Norton, D. F., & Norton, M. J. (1738). *A Treatise of Human Nature: Texts*. Clarendon Press.

Hutchins, E. (1995). *Cognition in the Wild* (Vol. 262082314). MIT press Cambridge, MA.

Hutchins, E. (2000). Distributed cognition. *International Encyclopedia of the Social and Behavioral Sciences*.

Hutto, D. (2013). Psychology's Inescapable Need for Conceptual Clarification. *A Wittgensteinian Perspective on the Use of Conceptual Analysis in Psychology*, 28.

Hutto, D. D. (2009). Lessons from Wittgenstein: Elucidating folk psychology. *New Ideas in Psychology*, *27*(2), 197–212.

Hutto, D. D. (2011). Enactivism: Why be radical. *Sehen Und Handeln*, 21–44.

Hutto, D. D., & Myin, E. (2013). *Radicalizing enactivism: Basic minds without content*. MIT Press.

Intons-Peterson, M. J., & Smyth, M. M. (1987). The anatomy of repertory memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(3), 490.

Intraub, H., Bender, R. S., & Mangels, J. A. (1992). Looking at pictures but remembering scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(1), 180.

James, W. (1890). *The Principles of Psychology*. Cosimo, Incorporated.

Johnson, M. K. (2007). Memory systems: A cognitive construct for analysis and synthesis. In H. L. Roediger, Y. Dudai, & S. M. Fitzpatrick (Eds.), *Science of memory: Concepts* (pp. 243–246). Oxford University Press.

Kandel, E. R. (2007). *In search of memory: The emergence of a new science of mind*. WW Norton & Company.

Kapur, S., Craik, F. I., Tulving, E., Wilson, A. A., Houle, S., & Brown, G. M. (1994). Neuroanatomical correlates of encoding in episodic memory: Levels of processing effect. *Proceedings of the National Academy of Sciences*, *91*(6), 2008–2011.

Kenny, A. (1963). *Action, emotion and will*. Routledge.

Kenny, A. (1973). *Wittgenstein* (Rev. ed.). Blackwell Publishing Ltd.

Kim, J. J., & Baxter, M. G. (2001). Multiple brain-memory systems: the whole does not equal the sum of its parts. *Trends in Neurosciences*, *24*(6), 324–330.

Kiverstein, J., & Clark, A. (2009). Introduction: Mind embodied, embedded, enacted: One church or many? *Topoi*, *28*(1), 1–7.

Klein, S. B., & Nichols, S. (2012). Memory and the sense of personal identity. *Mind*, *121*(483), 677–702.

Knowlton, B. J., Ramus, S. J., & Squire, L. R. (1992). Intact artificial grammar learning in amnesia: Dissociation of classification learning and explicit memory for specific instances. *Psychological Science*, *3*(3), 172–179.

Koriat, A. (2007). Remembering: Metacognitive Monitoring and Control Processes. In H. L. Roediger, Y. Dudai, & S. M. Fitzpatrick (Eds.), *Science of memory: Concepts* (pp. 243–246). Oxford University Press.

Kosslyn, S. M. (2005). Mental images and the brain. *Cognitive Neuropsychology*, *22*(3-4), 333–347.

Kosslyn, S. M., & Sussman, A. L. (1995). Roles of imagery in perception: Or, there is no such thing as immaculate perception.

Kryukov, V. I. (2008). The role of the hippocampus in long-term memory: is it memory store or comparator? *Journal of Integrative Neuroscience*, *7*(01), 117–184.

Lashley, K. S. (1929). *Brain mechanisms and intelligence: A quantitative study of injuries to the brain*. University of Chicago Press.

Lashley, K. S. (1950). In search of the engram. In *Symposia of the society for experimental biology* (Vol. 4, p. 30).

LeDoux, J. E. (2007). Consolidation: Challenging the traditional view. In H. L. Roediger III, Y. Dudai, & S. M. Fitzpatrick (Eds.), *Science of memory: Concepts* (pp. 171–176).

Lisman, J. E. (2007). Persistence: In search of molecular persistence. In H. L. Roediger III, Y. Dudai, & S. M. Fitzpatrick (Eds.), *Science of memory: concepts* (pp. 203–208). Oxford University Press.

Lisman, J. E., & Grace, A. A. (2005). The hippocampal-VTA loop: controlling the entry of information into long-term memory. *Neuron*, *46*(5), 703–713.

Loader, P. (2013). Is my Memory an Extended Notebook? *Review of Philosophy and Psychology*, *4*(1), 167–184.

Locke, J., & Winkler, K. (1689). *An Essay Concerning Human Understanding*. Hackett Publishing Company.

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior*, *13*(5), 585–589.

Logie, R. H., & Van Der Meulen, M. (2009). Fragmenting and integrating visuospatial working memory. In J. R. Brockmole (Ed.), *The visual world in memory* (pp. 1–32). Psychology Press.

Lowe, D. G. (1987). Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence*, *31*(3), 355–395.

MacNabb, D. G. C. (1962). *David Hume, A Treatise of Human Nature: Book 1*. Collins.

Malcolm, N. (1977). *Memory and Mind*. Cornell University Press.

Malenka, R. C., & Bear, M. F. (2004). LTP and LTD: an embarrassment of riches. *Neuron*, *44*(1), 5–21.

Marconi, D. (2012). Quine and Wittgenstein on the Science/Philosophy Divide. *Humana Mente Journal of Philosophical Studies*, *21*, 173–189.

Martin, C. B., & Deutscher, M. (1966). Remembering. *The Philosophical Review*, 161–196.

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

Matzel, L. D., & Miller, R. R. (2009). Parsing storage from retrieval in experimentally induced amnesia. *Learning & Memory*, *16*(11), 670–671.

Mayes, A. R. (2001). Aware and unaware memory: Does unaware memory underlie aware memory? In *Time and memory: Issues in philosophy and psychology*. Transactions of the Royal Society.

McClelland, J. L. (2000). Connectionist models of memory. *The Oxford Handbook of Memory*, 583–596.

McClelland, J. L., & Vallabha, G. (2007). Connectionist models of development: Mechanistic dynamical models with emergent dynamical properties. In J. P. Spencer, M. S. C. Thomas, & McClelland, James L. (Eds.), *Toward a unified theory of development: Connectionism and dynamic systems theory re-considered* (pp. 25–47).

McCormack, T. (2001). Attributing episodic memory to animals and children. *Time and Memory: Issues in Philosophy and Psychology*, 285–314.

McDaniel, M. A., & Masson, M. E. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(2), 371.

McGaugh, J. L. (2000). Memory--a century of consolidation. *Science*, *287*(5451), 248–251.

McGrath, M. (2007). Memory and epistemic conservatism. *Synthese*, *157*(1), 1–24.

McIntosh, A. R. (2007). Coding and representation: The importance of mesoscale dynamics. In H. L. Roediger III, Y. Dudai, & S. M. Fitzpatrick (Eds.), *Science of memory: concepts* (pp. 59–63). Oxford University Press.

Menary, R. (2010). Introduction to the special issue on 4E cognition. *Phenomenology and the Cognitive Sciences*, *9*(4), 459–463.

Merleau-Ponty, M. (1945). *Phenomenology of perception*. Motilal Banarsidass Publishe.

Meyers, C. D., & Waller, S. (2009). Psychological investigations: the private language argument and inferences in contemporary cognitive science. *Synthese*, *171*(1), 135–156.

Michaelian, K. (2011a). Generative memory. *Philosophical Psychology*, *24*(3), 323–342.

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

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

Milner, B. (1965). Visually-guided maze learning in man: Effects of bilateral hippocampal, bilateral frontal, and unilateral cerebral lesions. *Neuropsychologia*, *3*(4), 317–338.

Milner, B., Squire, L. R., & Kandel, E. R. (1998). Cognitive neuroscience and the study of memory. *Neuron*, *20*(3), 445–468.

Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychology Review*, *19*(3), 309–326.

Morgan, C. L. (1903). *An introduction to comparative psychology*. W. Scott, limited.

Morris, R. G., & Frey, U. (1997). Hippocampal synaptic plasticity: role in spatial learning or the automatic recording of attended experience? *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *352*(1360), 1489–1503.

Morris, R. G. M. (2007). Memory: Distinctions and dilemmas. *Science of Memory: Concepts*, 29–34.

Moscovitch, M. (2007). Memory: why the engram is elusive. In H. L. Roediger III, Y. Dudai, & S. M. Fitzpatrick (Eds.), *Science of memory: Concepts* (pp. 17–21). Oxford University Press.

Moyal-Sharrock, D. (Ed.). (2007). *Perspicuous Presentations: Essays on Wittgenstein's Philosophy of Psychology*. Palgrave Macmillan.

Moyal-Sharrock, D. (2009). Wittgenstein and the memory debate. *New Ideas in Psychology*, *27*(2), 213–227.

Moyal-Sharrock, D. (2013). Realism, but not empiricism: Wittgenstein versus Searle. *A Wittgensteinian Perspective on the Use of Conceptual Analysis in Psychology*, 153.

Nadel, L. (2007). Consolidation: The demise of the fixed trace. In H. Roediger, Y. Dudai, & S. M. Fitzpatrick (Eds.), *Science of memory: Concepts* (pp. 177–181). Oxford University Press.

Nadel, L., Samsonovich, A., Ryan, L., & Moscovitch, M. (2000). Multiple trace theory of human memory: computational, neuroimaging, and neuropsychological results. *Hippocampus*, *10*(4), 352–368.

Nader, K., & Einarsson, E. Ö. (2010). Memory reconsolidation: an update. *Annals of the New York Academy of Sciences*, *1191*(1), 27–41.

Nader, K., Schafe, G. E., & Le Doux, J. E. (2000). Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature*, *406*(6797), 722–726.

Nelson, K., & Fivush, R. (2004). The emergence of autobiographical memory: a social cultural developmental theory. *Psychological Review*, *111*(2), 486.

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

Nilsson, L.-G. (2000). Remembering actions and words. *The Oxford Handbook of Memory*, 137–148.

Noë, A. (2001). Experience and the active mind. *Synthese*, *129*(1), 41–60.

Noë, A. (2004). *Action in perception*. MIT Press.

Noë, A. (2009). *Out of our heads: Why you are not your brain, and other lessons from the biology of consciousness*. Macmillan.

Noë, A., Pessoa, L., & Thompson, E. (2000). Beyond the grand illusion: What change blindness really teaches us about vision. *Visual Cognition*, *7*(1-3), 93–106.

Noë, A., & Thompson, E. (2004). Are there neural correlates of consciousness? *Journal of Consciousness Studies*, *11*(1), 3–28.

Noice, H., & Noice, T. (2006). What studies of actors and acting can tell us about memory and cognitive functioning. *Current Directions in Psychological Science*, *15*(1), 14–18.

Noice, H., Noice, T., & Kennedy, C. (2000). Effects of enactment by professional actors at encoding and retrieval. *Memory*, *8*(6), 353–363.

Norman, K. A., Newman, E. L., & Detre, G. (2007). A neural network model of retrieval-induced forgetting. *Psychological Review*, *114*(4), 887.

O’Brien, G., & Opie, J. (2004). Notes toward a structuralist theory of mental representation. *Representation in Mind: New Approaches to Mental Representation*, 1–20.

O’Loughlin, I. (2011). The Persistence of “Storage”: Language and Concepts in Memory Research. In *European Perspectives on Cognitive Science*. New Bulgarian University Press.

O’Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, *24*(05), 939–973.

Orlandi, N. (2011). Embedded seeing: Vision in the natural world. *Nous 47*(4), 727-747.

Parfit, D. (1971). Personal identity. *The Philosophical Review*, 3–27.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*(12), 976–987.

Peterson, S. E. (2007). Learning: Multiplicity of mechanisms. In H. L. Roediger, Y. Dudai, & S. M. Fitzpatrick (Eds.), *Science of memory: Concepts* (pp. 49–52). Oxford University Press.

Piccinini, G. (2009). Computationalism in the philosophy of mind. *Philosophy Compass*, *4*(3), 515–532.

Piccinini, G. (2010a). The mind as neural software? Understanding functionalism, computationalism, and computational functionalism. *Philosophy and Phenomenological Research*, *81*(2), 269–311.

Piccinini, G. (2010b). The resilience of computationalism. *Philosophy of Science*, *77*(5), 852–861.

Pitt, D. (2013). Mental Representation. In *The Stanford Encyclopedia of Philosophy*. Retrieved from http://plato.stanford.edu/archives/fall2013/entries/mental-representation

Plato. (1990). *The Theaetetus of Plato*. (M. Burnyeat & J. Levett Trans.). Hackett.

Plaza, M., Gatignol, P., Leroy, M., & Duffau, H. (2009). Speaking without Broca's area after tumor resection. *Neurocase*, *15*(4), 294–310.

Polovina, J. J. (1984). Model of a coral reef ecosystem. *Coral Reefs*, *3*(1), 1–11.

Preston, J., & Schroeder, S. (2013). The Neuroscientific Case for a Representative Theory of Perception. *A Wittgensteinian Perspective on the Use of Conceptual Analysis in Psychology*, 253.

Prinz, J. (2006). Putting the brakes on enactive perception. *Psyche*, *12*(1), 1–19.

Prinz, J. J. (2006). Empiricism and State Space Semantics. In B. L. Keeley (Ed.), *Paul Churchland* (pp. 88–112). Cambridge University Press.

Prinz, J. J. (2012). *The conscious brain*. Oxford University Press.

Proudfoot, D. (1997). On Wittgenstein on cognitive science. *Philosophy*, *72*, 189–218.

Proust, M. (1925). *The Captive*. (C. K. Scott-Moncrieff & T. Kilmartin, Trans.). Modern Library.

Putnam, H. (1999). *The threefold cord: Mind, body, and world*. Columbia University Press.

Racine, T. P., & Slaney, K. L. (Eds.). (2013). *A Wittgensteinian Perspective on the Use of Conceptual Analysis in Psychology*. Palgrave Macmillan.

Radstone, S. (2008). Memory studies: For and against. *Memory Studies*, *1*(1), 31–39.

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

Randall, W. L. (2007). From Computer to Compost Rethinking Our Metaphors for Memory. *Theory & Psychology*, *17*(5), 611–633.

Reid, T. (1785). *Thomas Reid: Essays on the Intellectual Powers of Man; Edinburgh Edition*. Pennsylvania State University Press.

Roediger III, H. L. (1980). Memory metaphors in cognitive psychology. *Memory & Cognition*, *8*(3), 231–246.

Roediger III, H. L. (1996). Memory illusions. *Journal of Memory and Language*, *35*(2), 76–100.

Roediger III, H. L., Dudai, Y., & Fitzpatrick, S. M. (2007). *Science of memory: Concepts*. Oxford University Press New-York:

Roediger III, H. L., Gallo, D. A., & Geraci, L. (2002). Processing approaches to cognition: The impetus from the levels-of-processing framework. *Memory*, *10*(5-6), 319–332.

Roediger III, H. L., & McDermott, K. B. (2000). Distortions of memory. *The Oxford Handbook of Memory*, 149–162.

Roediger III, H. L., & Wertsch, J. V. (2008). Creating a new discipline of memory studies. *Memory Studies*, *1*(1), 9–22.

Rosen, D. A. (1975). An argument for the logical notion of a memory trace. *Philosophy of Science*, 1–10.

Rupert, R. D. (2004). Challenges to the hypothesis of extended cognition. *The Journal of Philosophy*, 389–428.

Russell, B. (1921). *The Analysis of Mind*. G. Allen & Unwin.

Ryle, G. (1949). *The concept of mind*. Routledge.

Sakaguchi, M., & Hayashi, Y. (2012). Catching the engram: strategies to examine the memory trace. *Mol Brain*, *5*, 32–32.

Sara, S. J. (2000). Retrieval and reconsolidation: toward a neurobiology of remembering. *Learning & Memory*, *7*(2), 73–84.

Sara, S. J. (2007). Consolidation: From hypothesis to paradigm to concept. In H. L. Roediger III, Y. Dudai, & S. M. Fitzpatrick (Eds.), *Science of memory: Concepts* (pp. 183–189).

Schacter, D. L. (2007). Memory: Delineating the core. In H. L. Roediger, Y. Dudai, & S. M. Fitzpatrick (Eds.), *Science of memory: Concepts* (pp. 23–27). Oxford University Press.

Schacter, D. L. (2008). *Searching for Memory: The Brain, The mind, and the past*. Basic Books.

Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., & Szpunar, K. K. (2012). The future of memory: remembering, imagining, and the brain. *Neuron*, *76*(4), 677–694.

Schacter, D. L., & Tulving, E. (1994). *Memory systems 1994*. MIT Press.

Schafe, G. E., Nadel, N. V., Sullivan, G. M., Harris, A., & LeDoux, J. E. (1999). Memory consolidation for contextual and auditory fear conditioning is dependent on protein synthesis, PKA, and MAP kinase. *Learning & Memory*, *6*(2), 97–110.

Schöner, G. (2009). Development as Change of System Dynamics: Stability, Instability, and Emergence. In J. P. Spencer, M. S. C. Thomas, & McClelland, James L. (Eds.), *Toward a unified theory of development: Connectionism and dynamic systems theory re-considered* (pp. 25–47).

Schwabe, L., Nader, K., & Pruessner, J. C. (2014). Reconsolidation of Human Memory: Brain Mechanisms and Clinical Relevance. *Biological Psychiatry*.

Schwitzgebel, E. (2002). A phenomenal, dispositional account of belief. *Nous*, *36*(2), 249–275.

Schwitzgebel, E. (2014). Belief. In *The Stanford Encyclopedia of Philosophy*. Retrieved from http://plato.stanford.edu/archives/spr2014/entries/belief/

Searle, J. (1998). The mind and education. In *Self awareness: Its nature and development* (pp. 3–11). New York, NY: Guilford Press.

Searle, J. (2007). Putting consciousness back in the brain. *Neuroscience and Philosophy: Brain, Mind & Language*, 97–124.

Searle, J. R. (1993a). The failures of computationalism. *Think (Tilburg, The Netherlands: Tilburg University Institute for Language Technology and Artificial Intelligence)*, *2*, 68–71.

Searle, J. R. (1993b). The problem of consciousness. *Consciousness and Cognition*, *2*(4), 310–319.

Sejnowski, T. J., & Rosenberg, C. R. (1987). Parallel networks that learn to pronounce English text. *Complex Systems*, *1*(1), 145–168.

Semon, R. (1925). The Mneme. *The Journal of Nervous and Mental Disease*, *62*(3), 332.

Senor, T. D. (2014). Epistemological problems of memory. In *Stanford Encyclopedia of Philosophy*. Retrieved from http://plato.stanford.edu/archives/spr2014/entries/memory-episprob/

Shanon, B. (1998). Metaphorical pluralism–not on the substantive level! *Behavioral and Brain Sciences*, *21*(01), 164–165.

Shapiro, L. (2012). What's New About Embodied Cognition? *Filosofia Unisinos*, *13*(2-suppl.), 214–224.

Shapiro, L. A. (2011). *Embodied Cognition*. Routledge.

Shea, N. (2007). Content and its vehicles in connectionist systems. *Mind & Language*, *22*(3), 246–269.

Shoemaker, S. S. (1959). Personal identity and memory. *The Journal of Philosophy*, 868–882.

Sluga, H. (1996). Whose house is that? Wittgenstein on the self. In *Cambridge Companion to Wittgenstein, ed. by Hans Sluga and David G. Stern* (pp. 320–353).

Smolensky, P. (1988). The constituent structure of connectionist mental states: A reply to Fodor and Pylyshyn. *The Southern Journal of Philosophy*, *26*(S1), 137–161.

Sorabji, R., & Aristote. (1972). *Aristotle on memory*. Duckworth London.

Spear, N. (2007). Retrieval: Properties and effects. *Roediger, Dudai, and Fitzpatrick*, *219*.

Spear, N. E. (1978). *The processing of memories: Forgetting and retention.* Lawrence Erlbaum.

Spencer, J. P., Perone, S., & Johnson, J. S. (2009). The dynamic field theory and embodied cognitive dynamics. In J. P. Spencer, M. S. C. Thomas, & McClelland, James L. (Eds.), *Toward a unified theory of development: Connectionism and dynamic systems theory re-considered* (pp. 86–118).

Spencer, J. P., Thomas, M. S., & McClelland, J. L. (Eds.). (2009). *Toward a unified theory of development: Connectionism and dynamic systems theory re-considered*. Oxford University Press New York, NY.

Squire, L. R. (2004). Memory systems of the brain: a brief history and current perspective. *Neurobiology of Learning and Memory*, *82*(3), 171–177.

Squire, L. R. (2006). Lost forever or temporarily misplaced? The long debate about the nature of memory impairment. *Learning & Memory*, *13*(5), 522–529.

Stanley, J., & Williamson, T. (2001). Knowing how. *The Journal of Philosophy*, 411–444.

Stern, D. G. (1991). Models of memory: Wittgenstein and cognitive science. *Philosophical Psychology*, *4*(2), 203–218.

Stern, D. G. (1995). *Wittgenstein on Mind and Language*. Oxford University Press.

Suddendorf, T., & Busby, J. (2003). Mental time travel in animals? *Trends in Cognitive Sciences*, *7*(9), 391–396.

Suddendorf, T., & Corballis, M. C. (1997). Mental time travel and the evolution of the human mind. *Genetic, Social, and General Psychology Monographs*, *123*(2), 133–167.

Susswein, N., & Racine, T. P. (2009). Wittgenstein and not-just-in-the-head cognition. *New Ideas in Psychology*, *27*(2), 184–196.

Sutton, J. (1998). *Philosophy and memory traces: Descartes to connectionism*. Cambridge Univ Press.

Sutton, J. (2007a). Batting, habit and memory: The embodied mind and the nature of skill. *Sport in Society*, *10*(5), 763–786.

Sutton, J. (2007b). Integrating the philosophy and psychology of memory: Two case studies. In *Cartographies of the Mind* (pp. 81–92). Springer.

Sutton, J. (2012). Memory. In *The Stanford Encyclopedia of Philosophy*. Retrieved from http://plato.stanford.edu/archives/win2012/entries/memory/

Sutton, J. (2014). Remembering as Public Practice: Wittgenstein, memory, and distributed cognitive ecologies. In D. Moyal-Sharrock, V. A. Munz, & A. Coliva (Eds.), *Mind, Language, and Action*.

Sutton, J., Harris, C. B., Keil, P. G., & Barnier, A. J. (2010). The psychology of memory, extended cognition, and socially distributed remembering. *Phenomenology and the Cognitive Sciences*, *9*(4), 521–560.

Ter Hark, M. (1995). Electric brain fields and memory traces: Wittgenstein and gestalt psychology. *Philosophical Investigations*, *18*(2), 113–138.

Ter Hark, M. (2000). Uncertainty, vagueness and psychological indeterminacy. *Synthese*, *124*(2), 193–220.

Thelen, E., & Smith, L. B. (1996). *A dynamic systems approach to the development of cognition and action*. MIT press.

Thompson, R. F. (2007). Persistence: Discrepancies between Behaviors and Brains. In H. L. Roediger III, Y. Dudai, & S. M. Fitzpatrick (Eds.), *Science of memory: Concepts* (pp. 199–202). Oxford University Press.

Thompson, W. L., Kosslyn, S. M., Sukel, K. E., & Alpert, N. M. (2001). Mental imagery of high-and low-resolution gratings activates area 17. *Neuroimage*, *14*(2), 454–464.

Thompson-Schill, S. L. (2003). Neuroimaging studies of semantic memory: inferring "how" from "where." *Neuropsychologia*, *41*(3), 280–292.

Toth, J. P. (2000). Nonconscious forms of human memory. *The Oxford Handbook of Memory*, 245–261.

Toth, J. P., & Hunt, R. R. (1999). Not one versus many, but zero versus any: structure and function in the context of the multiple memory systems debate.

Treves, A. (2007). Coding and representation: Time, space, history and beyond. In H. L. Roediger III, Y. Dudai, & S. M. Fitzpatrick (Eds.), *Science of memory: concepts* (pp. 55–58). Oxford University Press.

Trigg, J., & Kalish, M. (2011). Explaining how the mind works: On the relation between cognitive science and philosophy. *Topics in Cognitive Science*, *3*(2), 399–424.

Tsodyks, M. (2007). Coding and representation: on appealing beliefs and paucity of data. In H. L. Roediger III, Y. Dudai, & S. M. Fitzpatrick (Eds.), *Science of memory: Concepts* (pp. 69–73). Oxford University Press.

Tulving, E. (1983). Elements of episodic memory.

Tulving, E. (1991). Concepts of human memory. *Memory: Organization and Locus of Change*, 3–32.

Tulving, E. (1995). Organization of memory: Quo vadis. *The Cognitive Neurosciences*, 839–847.

Tulving, E. (2000). Concepts of memory. *The Oxford Handbook of Memory*, 33–43.

Tulving, E. (2005). *Episodic memory and autonoesis: Uniquely human*. New York: Oxford University Press.

Tulving, E. (2007). Coding and representation: searching for a home in the brain. *Coding and Representation: Searching for a Home in the Brain*, 65–68.

Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science*, *247*(4940), 301–306.

Tulving, E., Schacter, D. L., & Stark, H. A. (1982). Priming effects in word-fragment completion are independent of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*(4), 336.

Tversky, B., & Baratz, D. (1985). Memory for faces: Are caricatures better than photographs? *Memory & Cognition*, *13*(1), 45–49.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*(1), 231–270.

Van Gelder, T. (1991). What is the'D'in'PDP': a survey of the concept of distribution. *Philosophy and Connectionist Theory*, 33–59.

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

Varela, F., Thompson, E., & Rosch, E. (1991). *The Embodied Mind*. Cambridge: The MIT Press.

Vecht, C. J. (1989). Nociceptive nerve pain and neuropathic pain. *Pain*, *39*(2), 243–244.

Vosgerau, G. (2010). Memory and content. *Consciousness and Cognition*, *19*(3), 838–846.

Warrington, E. K., & Weiskrantz, L. (1968). A study of learning and retention in amnesic patients. *Neuropsychologia*, *6*(3), 283–291.

Waydo, S., Kraskov, A., Quiroga, R. Q., Fried, I., & Koch, C. (2006). Sparse representation in the human medial temporal lobe. *The Journal of Neuroscience*, *26*(40), 10232–10234.

Wearden, J. H., Denovan, L., & Haworth, R. (1997). Scalar timing in temporal generalization in humans with longer stimulus durations. *Journal of Experimental Psychology: Animal Behavior Processes*, *23*(4), 502.

Wegner, D. M. (1987). Transactive memory: A contemporary analysis of the group mind. In *Theories of group behavior* (pp. 185–208). Springer.

Weiskrantz, L. (1966). Experimental studies of amnesia. In C. Whitty & O. Zangwill (Eds.), *Amnesia* (pp. 1–31). London: Butterworths.

Wilbur, H. M., Tinkle, D. W., & Collins, J. P. (1974). Environmental certainty, trophic level, and resource availability in life history evolution. *American Naturalist*, 805–817.

Wilcox, S., & Katz, S. (1981). A direct realistic alternative to the traditional conception of memory. *Behaviorism*, *9*(2), 227–239.

Willmes, K., & Poeck, K. (1993). To what extent can aphasic syndromes be localized? *Brain*, *116*(6), 1527–1540.

Wilson, R. A. (2010). Extended vision. In N. Gangopadhyay, M. Madary, & F. Spicer (Eds.), *Perception, action and consciousness* (pp. 277–290). Oxford University Press.

Wittgenstein, L. (1965). *The Blue and Brown Books*. HarperCollins.

Wittgenstein, L. (1967). *Zettel*. University of California Pr.

Wittgenstein, L. (1972). *On Certainty*. HarperCollins.

Wittgenstein, L. (1974). *Philosophical Grammar: Part I, The Proposition, and Its Sense, Part II, On Logic and Mathematics*. University of California Press.

Wittgenstein, L. (1980). *Remarks on the Philosophy of Psychology, vol 1*, eds GEM Anscombe and GH von Wright, trans. GEM Anscombe. Blackwell.

Wittgenstein, L. (1982). *Last writings on the philosophy of psychology: preliminary studies for part ii of philosophical investigations*, eds H. Nyman & G. von Wright. University of Chicago Press.

Wittgenstein, L. (1993). *Philosophical Occasions, 1912-1951*. (J. C. Klagge & A. Nordmann, Eds.). Hackett Publishing.

Wittgenstein, L. (2010). *Philosophical investigations*. John Wiley & Sons.

Woll, S. (2001). *Everyday thinking: Memory, reasoning, and judgment in the real world*. Psychology Press.

Zahn, R., Moll, J., Paiva, M., Garrido, G., Krueger, F., Huey, E. D., & Grafman, J. (2009). The neural basis of human social values: evidence from functional MRI. *Cerebral Cortex*, *19*(2), 276–283.

Zaragoza, M. S., & Lane, S. M. (1994). Source misattributions and the suggestibility of eyewitness memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 934.