Summer 8-11-2020

Experience without Memory: Optogenetics, the Self, and the Ethics of Forgetting

David Kendall Casey
Georgia State University

EXPERIENCE WITHOUT MEMORY: OPTOGENETICS, THE SELF, AND THE ETHICS OF FORGETTING

by

DAVID KENDALL CASEY

Under the Direction of Neil Van Leeuwen, PhD

ABSTRACT

The horizon of clinical memory modification, long the domain of science fiction, is rapidly approaching; it is therefore imperative that we understand the ethical implications of such neuro-modificatory technologies. We might begin such inquiry with the public's worries about these technologies, namely that modifying memory will concomitantly modify the *self*. Yet, before discerning the reasonableness of this worry, we must understand the meaning of "the self" in relation to memory. Distilling this conception of the self is the principal aim of this thesis. I argue that many popular self-conceptions cannot capture our worries about neuromodification. Hence, I distill a novel such conception, which I call the *Proustian Self*—marshaling, to that end, not only neuroscientific evidence and metaphysical arguments but also literary-phenomenological analysis.

I ultimately argue that this conception should be the target of further neuroethical inquiry regarding the prospect memory modification and its effects on putative patients.

INDEX WORDS: Memory, Optogenetics, Memory modification, Self, Proustian self, Neuromodification, Neuroethics

EXPERIENCE WITHOUT MEMORY: OPTOGENETICS, THE SELF, AND THE ETHICS OF FORGETTING

by

DAVID KENDALL CASEY

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2020

EXPERIENCE WITHOUT MEMORY: OPTOGENETICS, THE SELF, AND THE ETHICS OF FORGETTING

by

DAVID KENDALL CASEY

Committee Chair: Neil Van Leeuwen

Committee: Marise Parent

Andrew Altman

Electronic Version Approved:

Office of Graduate Services

College of Arts and Sciences

Georgia State University

August 2020

DEDICATION

I dedicate this thesis firstly to my parents, but for whose unending support I completed this master's program; and secondly, to my two dogs, but for whose love on the hardest days I emerged from the program sane.

ACKNOWLEDGEMENTS

My foremost thanks and gratitude are to Professor Neil Van Leeuwen, who not only allowed this thesis the space to blossom and evolve on its own but also kept it focused and on track through its completion; and whose intellectual guidance, pedagogical example, and friendship I am fortunate to have garnered during my time at Georgia State University.

Secondly, I would like to thank the esteemed members of my thesis committee: Professors Andrew Altman and Marise Parent. I am fortunate and grateful to have received Professor Altman's ethical and political wisdom, not only by his contributions to my arguments herein but also by attending his life-changing seminar on race and racism during my very first semester. Likewise, I am fortunate and grateful for Dr. Parent, not only for the contributions she made to the following but also for furthering my intellectual and academic development in myriad ways; after taking classes in neurobiology and expressing to her my interest in scientific research, Dr. Parent took a chance on allowing this student of philosophy to work in her lab and assist with research that both informed my philosophical research and reoriented the trajectory of my academic career toward neuroscience and medicine.

I would also like to thank the entire department and faculty in the Department of Philosophy at Georgia State University, each of whom greatly contributed, either directly or indirectly—whether by teaching or mentoring me; or by contributing to such an esteemed and collegial department as it is—to my intellectual and personal growth.

Finally, I would like to thank those who set me on my journey into philosophy: the faculty in the Department of Philosophy at Rhode Island College, and especially Professor Matt Duncan, whom I am now delighted to call a *fellow* alum of Georgia State's philosophy master's program. Dr. Duncan's pedagogy and enthusiasm for philosophy undoubtedly spurred my interest in the

field, but it was his mentorship, guidance, and encouragement to pursue further study that led me here, to Georgia State, and to whatever lies beyond.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

By order of appearance:

OMM	Optogenetic memory modification
STM	Short-term memory
LTM	Long-term memory
SELF	Psychological self
NS	Narrative Self
AS	Autobiographical Self
PS	Proustian Self
EE	Episodic element
SE	Semantic element
ME	Mnemonic element

1 INTRODUCTION

> *Memory is a capricious and arbitrary creature. You never can tell what pebble she will pick up from the shore of life to keep among her treasures, or what inconspicuous flower of the field she will preserve as the symbol of "thoughts that do often lie too deep for tears." . . . And yet I do not doubt that the most important things are always the best remembered.*
>
> – Henry Van Dyke[1]

> *It is an old saying, that we forget nothing; as people in fever begin suddenly to talk the language of their infancy, we are stricken by memory sometimes, and old affections rush back on us as vivid as in the time when they were our daily talk. . . . Parting is death, at least as far as life is concerned. A passion comes to an end; it is carried off in a coffin or weeping in a post-chaise; it drops out of life one way or other, and the earth-clods close over it, and we see it no more. But it has been part of our souls, and it is eternal.*
>
> – W. M. Thackeray[2]

1.1 The fear of spotless minds

Americans are worried about their brains. In 2016, the Pew Research Center surveyed public attitudes about three emerging biotechnologies: prenatal gene editing; synthetic blood transfusion; and neural implants, ostensibly to improve "one's ability to concentrate and process information," which we'll call *neuromodification* (Funk et al., 2016). More Americans expressed concern about neuromodification than the other two technologies, with nearly three out of four predicting that these technologies "will become available before they have been fully tested or understood" (Funk et al., 2016). What explains this worry?

Not long ago, our autonomy, personality, and memory were thought to abide in an immaterial soul, undetermined by whatever might befall our frail and mortal bodies. But the more we've

[1] 1895, pp. 120–121.
[2] 1889, p. 194.

discovered about the neural underpinnings of these traits, the more deterministic, fragile, and mortal they have appeared.[3] Indeed, the Pew study's "foremost insight [was] that public wariness about emergent technologies is connected with concern over the potential loss of human agency," and that "proposals that would increase control over these technologies and keep humans in charge were met with greater acceptance" (Funk et al., 2018). In other words, the principal worry about neuromodification is that something might be done to us which we cannot control or supervise—in which we cannot participate.

Consider, as a manifestation of this anxiety, the film *Eternal Sunshine of the Spotless Mind* (Gondry, 2004). The plot centers around the relationship of odd-couple Joel Barish (played by Jim Carrey) and Clementine "Clem" Kruczynski (Kate Winslet): the former a socially inept, physically unkempt, and generally unremarkable forty-something; the latter an eclectic bookstore clerk, whose hair-color throughout the film changes as haphazardly as her moods, desires, and affections. After a particularly acrimonious (though, one gathers, not uncommon) breakup with Clem, Joel receives a letter from a mysterious firm called Lacuna, Inc., which informs him that Clem has sought Lacuna's services to erase all her memories of Joel—and that they've obliged. Unable to bear the grief, Joel decides that he, too, should undergo the procedure. The film then changes drastically, no longer moving forward chronologically but backward through Joel's memories of Clem, each of which he must re-experience before it is erased by Lacuna's technicians. In re-experiencing his memories, Joel realizes that he wants to keep them after all, that he would lose something important to himself—important to *the person he has become* through this relationship—should they disappear. Thus, Joel (more precisely, a mental avatar of Joel, for he remains

[3] Though cf. Nahmias (2011) for an argument that that a greater understanding of the physical determinates of consciousness, personality, agency, etc., results in more deterministic metaphysical beliefs only to the degree that one has faulty assumptions about what physicalism and determinism entail.

asleep during the procedure) spends the rest of the movie trying to hide his memories from the Lacuna technicians and salvage them from impending erasure.[4]

But why should audiences become so invested in Joel's plight? Why should we care whether he preserves his memories at all? Presumably, the impending erasure of Joel's memories doesn't *in itself* motivate our concern for him. For we, the audience, "knew" Joel before he attained those memories, and although we might have lamented that he was sad and lonely, we didn't lament that he didn't possess certain memories (or that he'd been, as far as we know, *dis*possessed of any). No, we're concerned for Joel because we believe that, after *losing* those memories, he will no longer be the same "Joel" that we've come to admire throughout the movie—throughout his relationship with Clem. With few exceptions, the worst fate that can befall a protagonist is death. But this memory erasure *is* a kind of death. Of course, "the human being called Joel Barish" doesn't die, but "the "Joel" who has, through his relationship with Clem, become so lively and complex, does, in a sense, *cease to exist* in losing that part of his mental life.

Indeed, this is why the story is so tragic.

The film ultimately reveals that Joel and Clem have been in a relationship *before*—and perhaps many times before—the one depicted in the film. And it is implied that they will meet

---

[4] The film's title comes from the poem "Eloisa to Abelard," by Alexander Pope (1717), penned as a fictional letter from the medieval French scholar Héloïse d'Argenteuil to her teacher, Pierre Abélard, whom she had secretly married. According to legend, when Héloïse's family discovered her affair with Abélard—she was almost twenty years his junior—they ordered him castrated. Abélard then entered the monastery and encouraged Héloïse to become a nun so that both of them could remain chaste, and thus eternally faithful to each other. Pope's poem depicts Héloïse deciding that, although still in love with Abélard, she would be better off not having to love in hopeless solitude, and that therefore, she would pray to forget their relationship—to bask in the "eternal sunshine" of resigned solitude (Nehring, 2005). In the fourteenth stanza, Héloïse proclaims:

> How happy is the blameless vestal's lot!
> The world forgetting, by the world forgot.
> Eternal sunshine of the spotless mind!
> Each pray'r accepted, and each wish resigned. (Pope, 1717)

again, only to erase each other, then meet again, and erase each other, and so on, *ad infinitum*. In the end, Joel's and Clem's "spotless minds" do not get to bask in the "eternal sunshine" of forgetfulness, but are instead condemned to a kind of Sisyphean nightmare: their endless cycle of love and erase resulting only in continual reversion to their former selves, thus squandering any prospect for self-improvement—which, in many ways, makes life worth living.

1.2 Memory modification and its discontents

Thus, *Eternal Sunshine* manifests some of our societal anxieties about neuromodification.[5] The worry, I think, is that we'll overestimate the promises of neuromodification and implement it too hastily, only to realize too late its irreparable consequences. And while the kind of precise, retroactive memory erasure depicted in *Eternal Sunshine* remains the domain of science-fiction, our concerns about neuromodification are not altogether unfounded. Indeed, neuroscientists have studied pharmacological memory modification since at least 1999 (Przybyslawski et al., 1999). Four years later, the President's Council on Bioethics cautioned researchers against attempting pharmacological memory modification, echoing the aforementioned worry about its potential effects on the self. "By 'rewriting' memories" with drugs, the Council warns, "we might succeed in easing real suffering at the risk of falsifying our perception of the world and undermining our true identity" (President's Council on Bioethics, 2010, p. 90).[6]

---

[5] *See* also the films: *Total Recall* (Verhoeven, 1990), *Inception* (Nolan, 2010), *Limitless* (Burger, 2011), *Total Recall* (Wiseman, 2012), *Lucy* (Besson, 2014), and *Transcendence* (Pfister, 2014), as well as multiple episodes of the science-fiction anthology series *Black Mirror* (prod. Jones & Brooker, 2011) and *Limitless* (prod. Sweeny, 2015)—the film's eponymous television spin-off—among many others I am surely forgetting.

[6] Original President's Council on Bioethics report (2003) excerpted in Farah (ed. 2010). Theoretical applications of pharmacological memory modification range from preventing drug relapse by modifying behavioral responses to environmental cues (Lee et al., 2005, 2006) to attenuating conditioned fear responses (Debiec & Nader, 2004), among multifarious others. Researchers usually perform these manipulations using drugs in the class *beta blockers*, which arrest the transmission of the stress-related neurotransmitter *norepinephrine*, typically binding to the postsynaptic beta-noradrenergic receptor. Pitman (2011) notes, however, that clinical applications in humans are not yet feasible.

Yet, one might wonder whether these worries arise from the proposed *means* of memory modification rather than memory modification *simpliciter*. That is, perhaps drugs are too imprecise to modify memories without risking deleterious off-target effects—but neuroscience has witnessed an explosion of ever more precise technologies: deep brain stimulation (DBS), for example, or transcranial magnetic stimulation (TCMS). Mightn't these technologies have fewer unintended effects? I don't think so; for memory is not merely an *aspect* of the self, as we normally conceive it, but rather *constitutes* the self; the two are so inextricably linked that any modification to the former—regardless of method—will modify the latter. This concern for the self, I contend, underlies our worries about neuromodification, and thus cries out for philosophical analysis.

1.3 Overview

Now, it may seem obvious that modifying memory will modify the self, but establishing this conclusion is not so simple. This is not because it is implausible; on the contrary, the connection between the self and memory is, as I've said, intuitive. Establishing this conclusion is difficult, rather, because while the connection between memory and the self is intuitive, precisely *how* they are connected is not. This ambiguity arises because the concepts "memory" and "self" are, themselves, ambiguous. In this thesis, I attempt to resolve this ambiguity. If we want to understand how memory modification might prove inimical to the self, we must first understand the science of memory and memory modification, along with what we mean by the "self." To this end, the following section (§2) concerns what *memory* is, how it works, and how it can plausibly be modified

with emerging neurotechnologies. Here, I use the nascent technology of *optogenetics* as a paradigmatic memory-modulatory mechanism.[7] Elucidating the *self*, then, is the purview of the third and fourth sections.

The self is a slippery concept. Indeed, the philosophical literature contains conceptions as multifarious as they are mutually incompatible. My target, however, is that conception of the self which is intuitively bound up with memory: that self with which, without memory (or insofar as memory has been modified), we would no longer identify—just as patients with amnesia or dementia, it often seems, are no longer the same personage that once inhabited their bodies. The primary aim of this paper is to distill this conception of the self. At least two criteria, I contend, must constrain this conception: a *uniqueness* criterion, which sets the self apart from all other selves such that each belongs to only one human person;[8] and a *diachronicity* criterion, which stipulates how the self persists through a lifetime.[9] The conception of self in which this analysis results, I call the *Proustian self*. I flesh out the Proustian self in the fourth section, first introducing the concept phenomenologically, then enumerating some of its essential qualities. In so doing, I draw heavily on Mark Rowlands's (2017) work on memory and the self.[10]

---

[7] Though I believe my argument can generalize, *mutatis mutandis*, to other neuromodifications in general, and to other types of memory modification in particular.

[8] Risking redundancy, I say "*human* person" here since I'm not in principle opposed to counting some animals as having (or *being*) selves, at least in the sense I have in mind. To keep things simple, however, we'll suppose herein that selves belong only to (or can only *be*) humans.

[9] This notion of diachronicity—viz., sameness over time—can subsume that the self can change gradually over the course of a life, just as the diachronicity of the physical-bodily subsumes that many of its components are not infrequently replaced. Thus, diachronicity can subsume a self that is drastically different from one time to another, but only as a result of each change leaving the better part of the previous self intact.

[10] Rowlands, too, conceptualizes a "literarily-named" self: the *Rilkean* self, after the German poet Rainer Maria Rilke. Although the Rilkean self is different from the Proustian self, the literary nomenclature is perhaps not entirely coincidental.

Aside from demystifying the relations between memory and the self, I intend for this argument to prepare the groundwork for further neuroethical analyses on (optogenetic) memory modification, in light of these newly clarified conceptions. While ethical analyses of memory modification often center on issues like autonomy, informed consent, or authenticity, I conclude by suggesting how we might include harms to the self among these concerns.

2 MEMORY, OPTOGENETICS, AND THERAPEUTIC FORGETTING

Joy's recollection is no longer joy, while sorrow's memory is a sorrow still.
– Lord Byron[11]

O, Memory, thou bitter-sweet—both a joy and a scourge.
– Germaine de Staël[12]

2.1 Memory formation and the hippocampus

Named after the Greek word for "seahorse," the hippocampus is a long, eggplant-shaped formation nestled within the brain's temporal lobes. Put simply, the hippocampus processes sensory and experiential information, mainly from the association cortices, and moves it into long-term memory.[13] I should note that in this section, when I refer to "memory," I mean *declarative* memory: for episodes as well as facts and meanings, anything that can be "declared"; as opposed to *non-declarative* memory: for skills, motor tasks, conditioning, and so forth.[14]

[11] *In:* Cantrevas et al. (Rudy, 2018).
[12] Ibid.
[13] The temporal lobes are the cortices on either side of the brain, just, as the name suggests, beneath the temples. In particular, the hippocampus, along with the amygdala and other structures, abides in the medial temporal lobes (MTL). The association cortices process and aggregate the various sensory inputs, for example: the visual cortex, the auditory cortex, the sensorimotor cortex, and so on
[14] It has been well established that the hippocampus is crucial for declarative memory, but not necessarily for the non-declarative variety. Nevertheless, since psychotherapy is often concerned with problematic events in one's life, and their associated (semantic) conceptual contents, it is plausible that therapeutic applications of memory-modulatory technology will be directed only toward declarative memory. My analysis is therefore directed at the same.

The hippocampus comprises what we might think of as a three-node circuit: each "node" representing a collection of neuronal cell bodies, each "wire" an axon jutting out toward the next node (*see* Fig. 2-1). These axons connect with each node at a junction called the *synapse*. The synapse is exceedingly complex, both morphologically and functionally, but in brief, it consists of a junction that facilitates communication between neurons. The junction itself consists of a pre-synaptic terminal, at the end of the axon, and a postsynaptic terminal, connected to a downstream neuron, which we are calling a node (or, more precisely, but one constituent of a node). So, when sensory information—touch, tastes, smells, sounds, and so forth; one might even include *thoughts* here—reaches the entorhinal cortex, a group of axons called the *perforant path* sends that information, in the form of an electrical signal, toward the circuit's first node, called the dentate gyrus (DG). At the end of the perforant path, this electrical signal gets converted into a chemical signal as it crosses the synaptic junction between the perforant path axons and their subsequent DG neurons; then, it's converted back to an electrical signal upon reaching the postsynaptic terminal—i.e., upon entering into the DG neurons (Davies & Morris, eds., 2004, chapters 4, 7; This process, called *synaptic transmission*, occurs at the junction of each "node" and "wire," and thus, we can take as concomitant to any neuronal communication referred to or implied herein.) The DG then relays the signal along its axons, the *mossy fibers*, to a region called the *cornu ammonis III*, (CA3).[15] It then travels along the CA3 neurons' axons, the *Schaffer collaterals*, ultimately reaching the CA1 region. This entire process is called *encoding*, and researchers theorize that it is during this stage, while information from one's sensory experience whirls around this tripartite circuit, that

[15] So named for its "horned" appearance.

one is said to be "holding information in one's head," such as a phone number or the eye color of a new acquaintance—otherwise known as *short-term memory*, or STM (Papanicolaou, 2006, p. 4).[16]

Not all information is sent from short- to long-term memory (LTM). After all, we needn't usually remember a phone number once we transfer it, as it were, from our heads onto the dial-pad. But the information our brain *does* deem worthy of remembering is sent back out of the hippocampus and into the neocortex, where it is put into LTM, ready to be recalled an hour, a day,

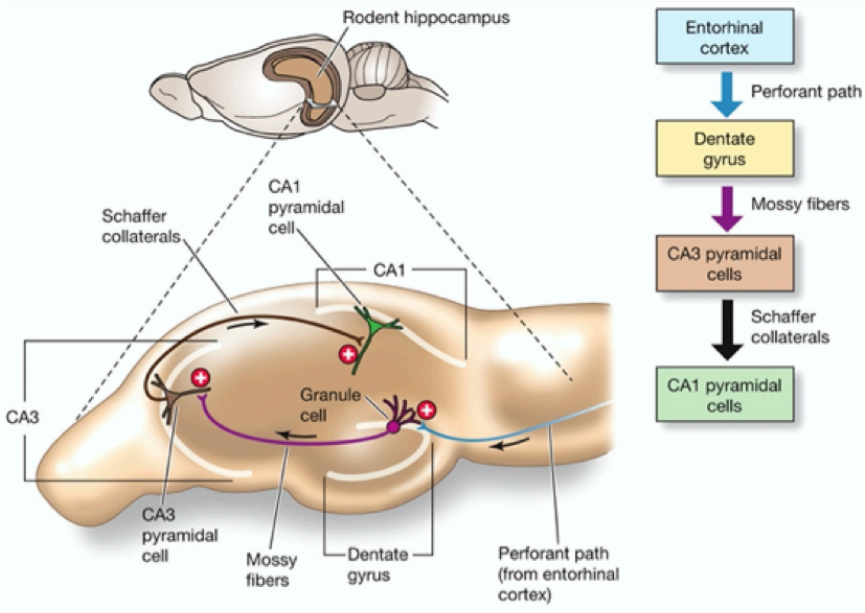

Figure 2-1 *The path of information through the rodent hippocampus. The entorhinal cortex sends axons comprising the perforant path to synapse on the DG. The DG then sends axons (mossy fibers) to the CA3, which, in turn, sends axons (Schaffer collaterals) to the CA1*. Reprinted from Rudy (2018, p. 25).

[16] Some readers might wonder whether this does not so much describe STM as it does *working memory*. While I don't want to get too deeply into the minutiae of this debate here, there exists some controversy over whether working memory is itself a *kind* of memory, like STM and LTM, or whether it is, rather, an *executive function* that makes use of the information in STM (*See*: Cowan, 2008). For particular reasons, I am partial to the latter notion; but whichever we accept is of no consequence (so far as I can see) to my present aims.

even many years later.[17] This process is called *consolidation*, and it is what most of us conceive when we think of "making a memory" (Rudy, 2018, pp. 288–311). When a memory is consolidated, it forms in the brain what neuroscientists call an *engram*, or *memory trace.* While we needn't know all the minutiae about engrams, this concept will return later (§4); so for now, we should at least get a basic understanding.

We can think of engrams as networks of neurons which represent (or "store" [though *see* note 17]) the memory of some experience. Reprising our earlier analogy, we might now imagine a circuit that not only sends electricity from node to node but that, thereupon, the connections between the nodes become stronger, or *potentiated* (we might imagine the wires becoming thicker). So, the more one rehearses, say, the digits of *pi*, the more one's brain sends information

[17] The term "storage," though useful in the abstract, is either an unfortunate vestige of how we used to conceive of long-term memory or a product of our currently impoverished conceptual vocabulary—or both—but the more we discover about how memory actually works, the less useful the word "storage" (as in, *computational* memory) becomes. As neuroscientist Jerry Rudy puts it,

> The content of the memory is contained in the unique *patterns of activity* in the neocortical regions of the brain activated by the experience. *There is no memory content* per se *contained in the hippocampus* [or, for that matter, the cortex]. All that it contains is the information that a specific pattern of activity in different cortical regions has occurred. This pattern is represented by (a) strengthening the synaptic connections between the input from the neocortex and the neurons activated in the hippocampus and (b) strengthening connections among the neurons that were activated. The only way the full content of the memory experience can be replayed is by outputs from the hippocampus activating the cortical representations of the experience. (2018, p. 311; my emphasis)

Note how Rudy's explication here, although closer to the mark, still uses conceptual language that implies that physical "storage" of some kind of content is involved in these neural processes. Despite stating, correctly, that no mnemonic *content* is stored in the hippocampus, Rudy nevertheless says that the hippocampus "contains . . . the information *that* . . . ," as if the hippocampus contains some kind of *propositional* information—the kind that usually, syntactically, follows "that" clauses, such as the information "*that* some pattern X inheres in hippocampal circuit Y." To me, this seems like another way of saying that the hippocampus "stores" or "contains" content, at least of the propositional variety, which seems undesirable for any theory that takes physicalism seriously. Thus, although still imperfect, I think that we should say content is, instead of "stored," *represented* in the brain, as a function of differential patterns of strengthened synapses and circuits (*see*, for instance: Tonegawa et al., 2018).

along those circuits wherein the memory of those digits has been consolidated. With each rehearsal, the connections within that circuit potentiate, resulting in ever easier communication between them. These neuronal circuits, composed of stronger connections relative to other such circuits, constitute individual engrams (*see* Fig. 2-2). In theory, the conscious experience of memory occurs when its underlying engram is activated (Levitan & Kaczmarek, 2015, pp. 493–527; Rudy, 2018, pp. 175–176).

Note that if one *prevents* a node within an electrical circuit either from receiving or transmitting a signal to the next node, then that circuit is broken and the information cannot travel to its destination. The same holds for engrams. So, to prevent a memory from consolidating, one must

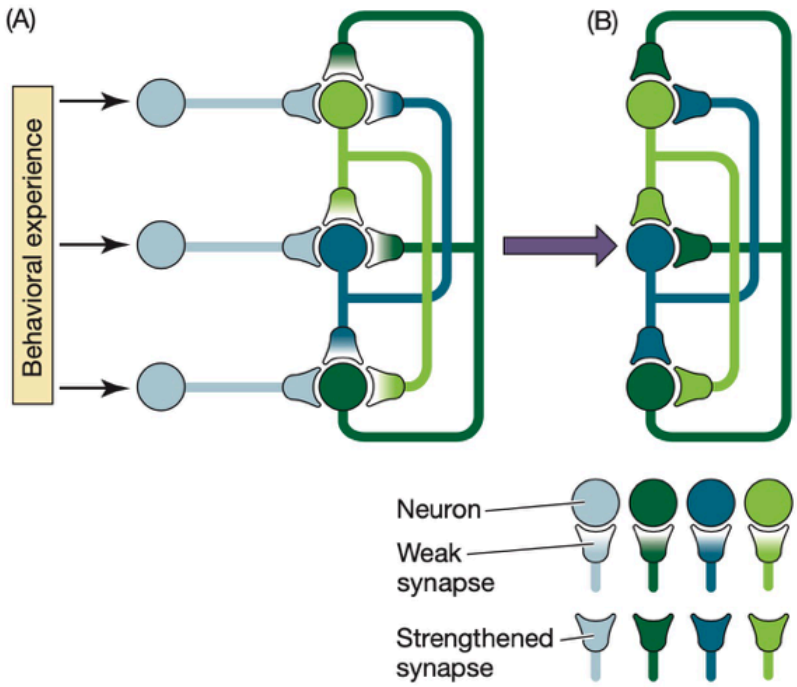

Figure 2-2 *The formation of an* engram *(or* memory trace*). For simplicity's sake, we might imagine that (A) sensory-experiential information is sent from the sensory cortices to the association cortices, into the hippocampus vie the perforant path, then into the* DG *(light blue neurons) and down the mossy fibers (light blue synapses). The mossy fibers then synapse onto neurons in the* CA3 *(dark green, dark blue, and light green neurons), whose synapses become stronger (B), thereby forming an engram designated to, or representative of, that experience. Thus, putatively, were the engram to become activated, a conscious memory of the original experience would ensue.* Reprinted from Rudy (2018, p. 176).

disrupt the relevant neuronal circuit, thereby preventing the connections between nodes from potentiating and, by extension, forming an engram (*see* Fig. 2-3). Now, the only question is how to disrupt the circuit.

2.2 Optogenetic memory modification

In brief, optogenetics consists of using light to excite or inhibit—activate or deactivate—cells, and thus allows researchers to manipulate neuronal activity with temporal, spatial, and functional precision. In other words, researchers can control not only the *duration* of neuronal excitation or inhibition but also which brain *areas* and which *kinds of neurons* in those areas to excite or inhibit (Fenno et al., 2011; Pedersen & Gross, 2018). As we've already noted, memories can consolidate only if the hippocampal circuit is complete. Therefore, if, during the period in which a memory is consolidating, a group of neurons is prohibited from receiving or transmitting a signal, then the information will not be able to leave the hippocampus and enter LTM. With optogenetic tools, researchers can throw a wrench, as it were, into the hippocampal circuit, thus blocking any consolidation that would otherwise have occurred.

To do this, researchers insert genetic material into the target population of neurons, causing them to express special proteins called *opsins*. When struck with light of a particular wavelength, opsins can activate other proteins (called *ion channels*) on those neurons' membranes, thereby exciting or inhibiting that particular neuronal population (Fenno et al., 2011, p. 391; *see* Fig. 2-4). The next step of optogenetic memory modification, naturally, is to insert a light source into the hippocampus. Typically, researchers surgically implant a fine glass tube, which, when attached to a light source—such as a laser or light-emitting diode (LED)—can direct light to areas as small as a single hippocampal field (although less-invasive strategies have also been theorized [*see* Fig. 2-

4]). Researchers can then shine light on the group of hippocampal neurons under study, thereby disrupting any signal coming into that region and "breaking" the circuit (*see* Fig. 2-3).

Optogenetics, broadly speaking, has myriad applications both within and outside of the brain.[18] But to what end is optogenetic *memory* modification?

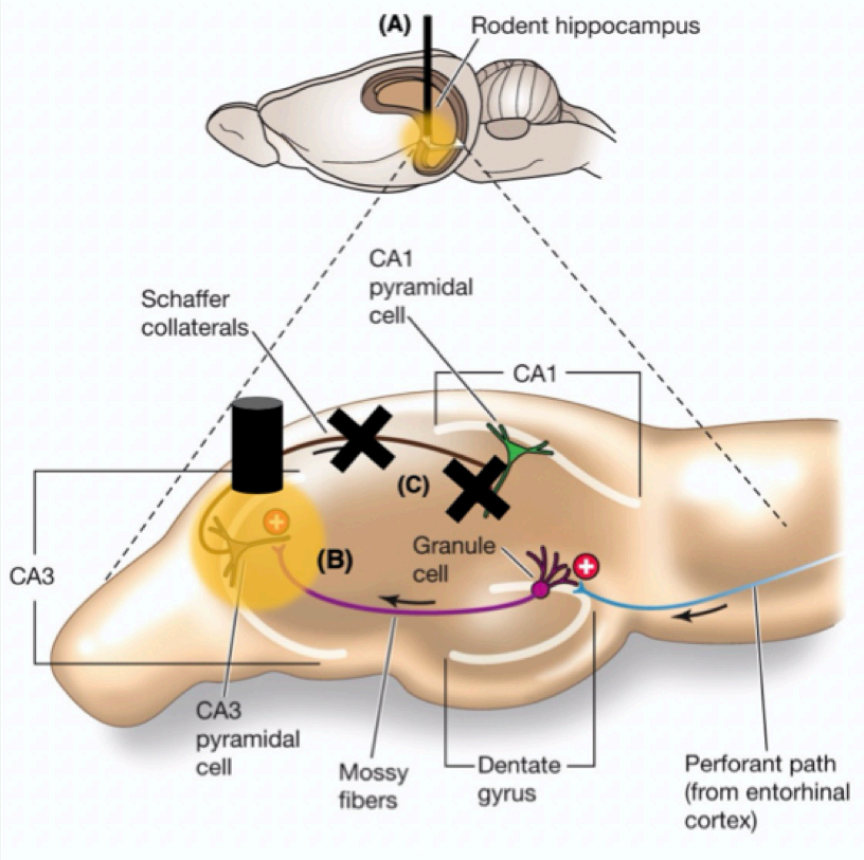

Figure 2-3 *A rodent hippocampus—stipulatively expressing inhibitory opsins in CA3 pyramidal neurons—undergoing optical stimulation. (A) A fine glass tube is surgically inserted into the relevant brain area, then (B) light from a laser or LED is shone through the tube, onto that region, which (C) inhibits the neurons comprising that region, and, in turn, stops information from propagating along the rest of the circuit.* Reprinted from Rudy (2018, p. 25), with alterations.

[18] *See*, for instance: Bingen et al. (2014); Busskamp et al. (2012); Cela et al. (2019); Crocini et al. (2016); DiGuiseppi and Zuo (2019); Feliu et al. (2018); Gaub et al. (2015); Huang et al. (2012); Kravitz et al. (2010); Moser (2015); Pama et al. (2013); Paralikar et al. (2010); Pedersen and Gross (2018); and Touriño et al. (2013).

2.2.1 Memory and eating

Of all the purposes for which humans evolved conscious memory, *knowing when and how much to eat* seems an unlikely candidate. Our drive to eat, we might think, is so evolutionarily primitive, so fundamental, that we should never have needed any mechanism more complex than basic instinct to regulate it. Yet, neuroscientists Marise Parent's lab has demonstrated precisely the contrary. Parent and her team were not the first to hypothesize a relation between memory and eating, but, for the last decade or so, have nevertheless been at the vanguard of such research (Hannapel et al., 2017; Henderson et al., 2013; Parent, 2016b, 2016a; Parent et al., 2014). And in 2019, they became the first to use optogenetics to demonstrate the link between the two.

Their experiment ran as follows. Once the relevant—in this case, *inhibitory*—opsin was virally induced and optical tubes attached to lasers were inserted into the hippocampi of experimental rats, they were fed a meal. Because the researchers could activate the laser at any time, they decided to test three activation points: First, they would activate it (thereby inhibiting the hippocampus), ten minutes *before* the first daily meal; for another group, they'd inhibit it *during* the first ten minutes of the meal; and for another, for ten minutes, five minutes *after* the end of the meal. "The goal," the authors state,

> was to identify when neural activity in these [hippocampal] neurons is necessary for inhibiting intake and to determine whether neural inhibition given for 10 min after the first sucrose meal, when the memory of the meal would be undergoing consolidation, would promote the initiation of the next meal and increase intake during that next meal when the neurons were no longer inhibited. (Hannapel et al., 2019, p. 7)

They wanted to know, in other words, whether making the rats forget that they ate would cause them to eat sooner, and more, than they would have otherwise.

They found that when hippocampal neurons were inhibited for the first ten minutes *after* the end of the meal, when the memory was likely consolidating, the rats indeed ate larger subsequent meals, and ate them sooner, than they would have otherwise. Moreover, anticipating the objection that hippocampal inhibition might affect interoceptive hunger signals (and that it is therefore not the lack of memory *per se* that dysregulates eating), Parent and her lab also inhibited rats' hippocampi after a meal consisting only of *saccharin*, a chemical sweetener that does not modulate hormonal signals for hunger or satiety. They got the same result. "The finding that postmeal optogenetic inhibition increased future saccharin intake," they note, "suggests that the ability of [hippocampal] neurons to control future intake does not require post[-meal] interoceptive visceral signals"—that *memory alone* can regulate one's eating (Hannapel et al., 2019, p. 11).

This research not only helps explain observed correlations between weight gain and memory deficits (e.g., Cheke et al., 2016; Martin et al., 2018) but also portends many clinical applications. For example, the researchers suggest that, as a corollary to their findings, "enhancing the memory of a meal may be a promising strategy for limiting intake and promoting weight loss" (Hannapel et al., 2019, p. 13). Their results certainly suggest this possibility. But what if we didn't have to infer such a corollary to glean clinical applications from this research? What if, that is, we could use the findings that forgetting a meal will increase eating, not to help people eat *less*, but to eat *more*?

2.2.2 Clinical applications

Consider the eating disorder *anorexia nervosa* (henceforth, simply "anorexia"). In her memoir, *Elena Vanishing*, anorexia survivor Elena Dunkle vividly describes her state of mind when it came to eating. She recalls a nurse's aide once presenting her with a tray of hospital food: "No," she thinks. "There's no way I can force that stuff down. This morning, I had three bites of pudding,

and I'm still full. At the thought of more food, the familiar pains knife through me" (Dunkle & Dunkle, 2015, p. 1). She recalls another occasion when, trying to placate her worried mother, she forced down a cup of chocolate pudding on pain of relentless self-harassment: "*Twenty-five grams of simple carbs*," her inner voice jeers. "*Twenty-five grams at least! Insulin is flooding your bloodstream right this minute, turning sugar molecules into fat!*" (Dunkle & Dunkle, 2015, p. 29; author's italics). June Alexander, an Australian anorexia sufferer from age eleven into her fifties, likewise had what she calls "an intense fear of eating," adding, "if I did eat, the guilt would be totally enormous, and I would have to go and do some physical activity to compensate" (Arnold, 2013, p. 10). While we cannot reduce anorexia's causes and contributors to a single factor,[19] these phenomenological accounts (among others) reveal one cognitive component that researchers and clinicians often overlook: *memory*.

The language these survivors use hints at the role memory plays in their disorders. When Dunkle is brought a tray of food, for example, she does not merely think she is not currently hungry or has no appetite; she claims that she is *full* from eating, hours earlier, a mere three bites of pudding—which by this point would have left her stomach. In other words, she feels full (at least partly) because *remembers* her previous meal. Her memory also provides ammunition for her critical inner voice. Most people who eat a meal do not proceed to ruminate on its molecular metabolism (even if, perhaps because watching one's weight, one ruminates on one's "metabolism" as an abstract concept). But Dunkle's anorexia made her unlike most people. She could not stop thinking about her food accumulating in fat, her mind spurred on by the memory of having consumed it. Similarly, although Alexander describes a fear of eating itself—a fear when

[19] Although research continues to reveal new insights. For two excellent and comprehensive works on the state of the art of eating disorder research, *see*: Arnold (2013) and Giordano (2005). For a comprehensive history of both the populace's and the scientific and medical communities' understanding of the disorder, *see*: Brumberg (2000).

looking forward to the activity of eating; a *prospective* fear—her anguish seems partly attributable to her memory of eating, to the "guilt" she feels, retrospectively, about the fact that she ate. We can adduce these considerations, along with Parent's research, to the claim that not only are eating and memory intertwined but eating *disorders* and memory likely are as well.

Should optogenetic memory modification (herein: OMM) one day become applicable to humans, psychotherapists will almost certainly want to use it to treat cognitive-behavioral disorders that implicate memory. In 2019, for instance, a team of Canadian researchers found that optogenetically activating negative memories in rats increased their depressive symptoms, in light of which they suggested that "[i]nhibiting negative memory engrams in the hippocampus could be a novel therapeutic approach for treating cognitive symptoms," such as rumination on negative memories, "in depression" (Zhang et al., 2019, p. 7588). We might expect that researchers and clinicians will want to do the same for eating disorders as the connection between memory and eating becomes clearer, and OMM more feasible.

Imagine, for example, that a severely anorexic patient—call her Annie—enters treatment. Annie is not so unhealthy that she needs immediate nutrition, but she is quickly approaching that point. So, drawing on research linking memory deficits to increased eating, her physician reaches out to a biotechnologist and requests that she develop a small, portable optogenetic implant, which we'll call an *optrode*—similar, perhaps, to the electrodes currently used for deep-brain stimulation (*see* also: Fig. 2-4). The biotechnologist soon delivers the requested device to Annie's physician, who hastens to implant it in her hippocampus. Now, Annie is no longer guilt-ridden after she's eaten, nor excoriated by the voice in her head, nor tormented by phantasms of metabolizing food; for after she eats, she or her doctor need only press a button and her optrode blocks the memory

of the meal from forming. As a result, Annie begins to eat more, and more frequently, than she would have otherwise.

At this point, one might ask: first, if OMM is not yet feasible in humans, why worry about it? And second, if we have not yet established the link between memory and eating in humans, why believe OMM would work the same as it does in rats?

Regarding the first question: Certainly *some* neuroscientific speculation too fantastical to be worthy of analysis. In fact, Gilbert, Harris, and Kapsa (2014) hold this view regarding optogenetics. "It could be speculated that *if* neuroscientists can control neurons in mice," the authors state,

> *then* in the near future, one could control the mind of a human; *therefore* we ought to explore the ethical questions relating to this mind control possibility in humans. Such speculation detracts from focusing on more immediate ethical needs related to application[s] such as optogenetic restoration of vision. (Gilbert et al., 2014, p. 4; author's italics)

But this is a red herring. Gilbert et al. hastily equivocate controlling *neurons* with "mind control," but I am not concerned here with the possibility of mind control—and I'd be surprised if anyone really were. Rather, the present speculation is warranted by a rapidly expanding body of research suggesting not only that optogenetic treatments in humans are feasible, but also that many researchers are interested in, and actively working on, realizing such applications.[20] Typically, biotechnology advances from trials on rodents, such as rats, to primates, such as chimps or macaques, then to humans. In that respect, optogenetics has begun to round second base: Researchers have

[20] *See* above (p. 14). *See* also: West (2014) for an interview with two of the foremost current optogenetics researchers, Xu Liu and Steve Ramirez. While they don't furnish any particular empirical evidence of the feasibility of optogenetic therapy in humans, their opinion is nevertheless more valuable than lay-speculation. When asked whether they pictured optogenetic memory modification "being used in humans in, say, 10 years [by 2024]," Ramirez says: "In principle, [it's] possible, but the technology doesn't exist yet. I'd love to see it implemented and it's just one revolution away from happening, but currently, I don't think it's possible. . . . The technology isn't there yet." (West, 2014).

studied, and continue to study, optogenetic methods in non-human primates, and have even verified the safety and effectivity of the viral vectors which carry gene-promoters for opsin expression (adeno-associated virus, or AAV), along with the opsin genes themselves, in these primates, reasoning, in turn, that either AAV or a variant of it will be safe for humans (Deng et al., 2018; Diester

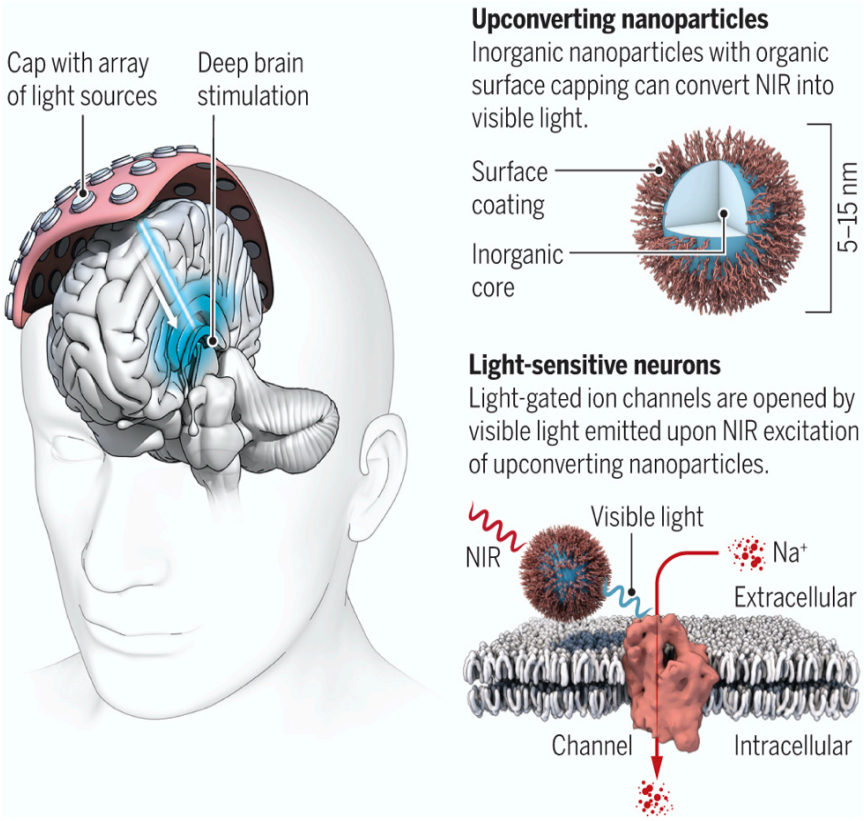

Figure 2-4 *A proposal for a non-invasive mechanism of optogenetic neuromodification. The patient wears a cap dotted with light sources that emit invisible, near-infrared (NIR) light, which can permeate the skull and brain matter. Once the NIR reaches its target brain region, it collides with a nanoparticle expressed by genetically modified neurons in the target region. The nanoparticle's surface coating transforms the NIR into visible light which can activate an opsin in the target neuron's membrane, which in turn activates a particular ion channel, thereby activating (exciting or inhibiting) the neuron.* Reprinted from Feliu et al. (2018); graphic by C. Bickle/*Science*.

et al., 2011).[21] So while I will grant that optogenetic therapy in humans will require much more research, nothing suggests it is in principle impossible. Indeed, all current evidence suggests its plausibility. It is therefore prudent to examine the philosophical issues surrounding OMM's potential applications.

As for the second worry, while studies on rodents or primates do not always generalize to humans, I think we have good reason to believe that Parent et al.'s study will. First, as the researchers state:

> Excess intake of fats and/or sugars and obesity in rodents impair hippocampal synaptic plasticity and hippocampal-dependent memory. Hippocampal dysfunction, in turn, increases meal frequency and food intake. In humans, being overweight . . . is associated with hippocampal atrophy and episodic memory deficits, and enhancing the memory of a meal may be a promising strategy for limiting intake and promoting weight loss. (Hannapel et al., 2019, p. 13; see original for citations)

We also know that disrupting post-meal memory in humans causes them to eat more during the next meal, and that enhancing it causes them to eat less (Higgs, 2016; Robinson et al., 2013). Moreover, patients with *anterograde amnesia*, the inability to make new memories, will eat to satiety, then eat *again* when presented with another meal, despite presumably being full (Hebben et al., 1985; Higgs, 2008; Rozin et al., 1998). The converse relation seems to obtain as well: Deficits in episodic memory correlate with uncontrolled eating and increased weight (Cheke et al., 2016; Martin et al., 2018). This is all to say that the relation between memory and eating in humans has been well established.[22]

---

[21] If one should wonder whether opsin expression is only analyzable *in vitro* and is therefore not feasible for human use, I point out that many of these same researchers have pioneered methods for analyzing opsin expression *in vivo* in non-human primates (Deng et al., 2018; Diester et al., 2011; Yazdan-Shahmorad et al., 2018).

[22] Humans are of course better cognitively endowed than laboratory rats, so even if a patient does elect to undergo this kind of therapy, we might imagine that she can simply infer that she's eaten recently, even when she doesn't remember doing so. But I don't think this makes OMM a less attractive prospect for treating these disorders. As adduced by Dunkle's and Alexander's accounts, memory seems to play an important role in at least sustaining, if not also initiating, the negative thought patterns that can perpetuate dysfunctional eating behavior. In the hospital, Dunkle doesn't

In sum, if optogenetic inhibition can be applied to humans—for which I've given reasons to believe—then it could plausibly realize the same cognitive-behavioral effects on eating as it does in laboratory experiments, given the relationship between memory and eating of which we already know. Nevertheless, OMM has yet to be tried in humans, so I can at best only *assume* here that it one day will, and that it will be able to modulate human eating behavior. Could both turn out to be false? Of course. The evidence nevertheless suggests that they're plausible enough to warrant the present analysis.

3 IN PURSUIT OF THE SELF

At the risk of sounding trivial: What *good* is memory? Sure, we might conjure up endless Darwinian "just so" stories about memory's *purpose*: maybe it allowed our ancestors to remember where they could find food, or with which members of their group they were on bad terms. But I'm asking a deeper question about memory, perhaps better rephrased as: What would we be *without* it? Consider patients with severe anterograde amnesia, such as Clive Wearing, a renowned musicologist, who, at the apogee of his career, contracted a neurological disease that left him unable to make new memories. The neurologist Oliver Sacks once wrote of Wearing:

> [His] journal entries consisted, essentially, of the statements "I am awake" or "I am conscious," entered again and again every few minutes. He would write "2:10 PM: This time properly awake. . . . 2:14 PM: this time finally awake. . . . 2:35 PM: this time completely awake," along with negations of these statements: "At 9:40 PM I awoke for the first time, despite my previous claims." This in turn was crossed out,

want to eat because she remembers eating three bites of pudding many hours earlier, and when she does eat, she's haunted by the thought of that food converting to fat. Likewise, Alexander's own worst enemy was the guilt she felt after eating, guilt that would drive all manner of compensatory behaviors, such as intense exercise (which is, of course, a healthy activity in normal circumstances, but is, for anorexic patients, who are often malnourished, dehydrated, and have low energy stores and hormonal imbalances, probably not). Thus, the OMM-induced instinct to eat more, as observed in Parent's study, may not even ameliorate dysfunctional eating, in humans at least, as much as will ostensibly eliminate the guilt, rumination, and other conscious feelings about having eaten (though one can assume the former will no doubt play some, perhaps largely unconscious, role).

> followed by "I was fully conscious at 10:35 PM, and awake for the first time in many, many weeks." This in turn was cancelled out by the next entry. (2007)

Wearing's experience sounds abysmal to those of use with normal mnemonic capacities.[23] Indeed, that conscious experience could be so mangled by the loss of memory likely motivates, at least in part, our misgivings about trifling with our brains.

Yet, on reflection, this worry seems puzzling; for most of us cannot remember *most* of our lives. Indeed, most of us cannot recall much of the previous day—save for a few evanescent scenes, sometimes fused together to appear continuous. But if we're fine with forgetting whole days, weeks, months, even *years* of our lives, why do we dread forgetting a little more? Why might we dread being like Clive Wearing, if only for the memories surrounding the manipulation? I think this discrepancy reveals that we take the ordinary phenomenon of failing to remember to be vastly different from disrupting memory consolidation, whether neuropathologically or via neuromodification.

The intuition, it seems, is that we simply *are* what we (naturally) *remember*—that, somehow, "Clive Wearing: the shrewd musicologist," is not just hidden from view, tucked away under the lesions left by his disease, but that he *ceases to exist*, that he disappeared along with his memory, and was replaced with a kind of glitchy simulacrum.[24] In short, we seem to intuitively

---

[23] Appropriately, Sacks's article, which he wrote for the *New Yorker*, is titled "The Abyss" (2007).

[24] Apparently, Wearing's memory only lasted about thirty seconds. It simply defies common sense to think that someone with such an impoverished memory could maintain an integral self-conception, however we want to understand that notion at this point. For example, Sacks, giving harrowing insight into Wearing's (likely) phenomenology, writes:

> Though one cannot have direct knowledge of one's own amnesia, there may be ways to infer it: from the expressions on people's faces when one has repeated something half a dozen times; when one looks down at one's coffee cup and finds that it is empty. . . . Lacking memory, lacking direct experiential knowledge, amnesiacs have to make hypotheses and inferences, and they usually make plausible ones. . . . Yet Clive, rather than making plausible guesses, always came to the conclusion that he had just been "awakened," that he had been "dead." This seemed to me a reflection of the almost *instantaneous effacement of perception* for Clive—*thought itself* was impossible within this

believe that the self is, either wholly or mostly, constituted by memory. And if this is the case, then the concern about memory modification is, entirely or mostly, motivated by the intuition that we might *lose our selves* in the process. This claim seems plausible. But before we can establish it, we must clarify what we mean by the *self*. This conception should capture our intuitions about cases like Wearing's—that the loss of self (if not in whole, then in some integral part) will attend any *loss* of memory, and vice versa. In other words, our conception of the self must relate to (be linked to, bound up with, etc.) memory, such that modifying the latter will necessarily modify the former. In this section, I survey a number of conceptions of the self, ultimately denying each in turn as the conception that we are interested in, thereby bringing the conception we *are* interested in into sharp relief.

3.1 The Self in Question

There are many ways we might conceive of the self. One might think of the self, in a basic, primitive sense, as a *locus of experience*: the "center," as it were, of everything we perceive, including our sensations, thoughts, beliefs, emotions, desires, and memories. We'll call this conception of the self the *metaphysical* self.[25] Note that the metaphysical self is a *passive* self, less the director of our conscious experience than its audience. Yet, this minimal conception doesn't capture what most people mean by "the self," and is thus not the conception we're after. The self is generally considered unique: that, and only that, which answers to the name of "I," and which others would recognize as essentially *me*. But metaphysical selves are not unique in this way. A mere locus of

> tiny window of time. Indeed, Clive once said to [his wife] Deborah, "I am completely incapable of thinking." (2007; my emphasis)

[25] Galen Strawson and Dan Zahavi, respectively, call this self the *minimal subject* or the *experiential self* (Strawson, 2009, 2011; Zahavi, 2011).

experience could be divested of every cognitive ability save for the dimmest light of consciousness (that of an ant, say, or Chalmers's thermostat[26]) and remain the *same* self; or, conversely, could switch places with anyone else's, and as long as the higher-level cognitions and conations remained, nothing would seem untoward to their subject.[27] Thus, we cannot be referring to this conception when we consider Wearing's loss of "self."

Luckily, we needn't only think of the self in this sense. Indeed, many more robust conceptions exist in the philosophical and cognitive-scientific literature. These, I will call *psychological* selves. On the psychological conception, the self comprises more than our bare experience, and is often constituted by a manifold of thoughts, beliefs, desires, intentions, and memories, among other cognitive and conative phenomena. Thus, as opposed to metaphysical self-conceptions, these conceptions view the self as unique to each individual human who owns (or *is*) one. I should note that, since we're most interested here in how the self is constituted by *memory*, we can ignore for

---

[26] *See* Chalmers (1996, pp. 293–299): Chalmers uses the hypothetical example of a conscious thermostat—"an information-processing system [like the physical brain] that is almost maximally simple" (p. 293)—to draw out our intuitions of wherein the most basic conscious experience would consist: "Certainly it will not be very interesting to be a thermostat," he says. "The information processing is so simple that we should expect the corresponding phenomenal states to be equally simple" (p. 293). "To make this view seem less crazy," he continues,

> we can think about what might happen to experience as we move down the scale of complexity. We start with the familiar case of humans, in which very complex information-processing gives rise to our familiar complex experiences. Moving to less complex systems, there does not seem much reason to doubt that dogs are conscious, or even that mice are. (p. 294)

He thus moves down the scale—"through lizards and fish to slugs," then to "simple neural networks all the way to thermostats"—asking, finally, "where should consciousness wink out?" (pp. 294, 295). Whatever we think about the metaphysical reality of this scaling (whether, say, we assert that it's a Sorites paradox and is therefore metaphysically specious), we should at least countenance some "maximally simple" consciousness in the *human* case. This is what I am here calling the *metaphysical* self.

[27] "Anyone" here meaning, any other biological human being *simpliciter*; I'm constricted here to using intuitive language, and am not in so doing trying to beg the question.

now how other cognitions like belief, desire, and emotion figure into it. These would be the purview of a separate project.

In what follows, I will first survey the different species of psychological selves, painting only in broad strokes, since examining the minor variations between each is would take us too far afield of my current project.[28] To give a preview: The following conceptions ultimately fail to capture the concomitance of memory and the self; but in so failing, as if carving away a mold, they bring into relief the conception that ultimately *succeeds*, and which will be the object of §4.

3.2 The Psychological Self

The *psychological* self is the notion most in line with our everyday, intuitive concept of selfhood. As far back as the 17th century, for instance, John Locke defined the self as:

> that conscious thinking thing, whatever substance made up of (whether spiritual or material, simple or compounded, it matters not), which is sensible, or conscious of pleasure and pain, capable of happiness or misery, and so is concerned for itself, as far as that consciousness extends. (1979, p. 346)

I imagine that most people could not give a better, if no less vague, definition. Numerous philosophers after Locke have given equally numerous, often mutually incompatible, definitions of the self—including those, such as Hume and his followers, along with many in Eastern traditions, who claim that the entity we often call "the self" is an illusion (Hume, 1888; Siderits, 2011).[29]

While these latter views should not be dismissed offhand, we can in fact remain metaphysically agnostic, for my purposes, about the psychological self and its subtypes. That is, we needn't

---

[28] For a wide-ranging volume on the topic, *see*: Gallagher (2011).

[29] Hume is often considered the progenitor of "bundle theory," which, in his words, maintains that there is no such entity as a "simple and continu'd" self, but rather, there are only "the successive perceptions . . . that constitute the mind." "[N]or," he continues, "have we the most distant notion of the place where these scenes are represented, or of the materials, of which it is comprised" (Hume, 1888, p. 253). In other words, Hume doesn't think there exists a self that persists through time or bears witness to—or directs—our thoughts, emotions, beliefs, and actions. He thinks, rather, that there is only time, there are only our thoughts, emotions, beliefs, and actions, but nothing behind (or in front of) them, "holding them together" as it were, in one person or through a period of time.

determine *whence* the self in question, or whether it *exists*, metaphysically. Why not? Because even if our target is a purely phenomenological construct, we are primarily concerned with changing (or eliminating) that construct upon modifying its neural underpinnings. In other words, we can draw ethical conclusions from subjective experience, independent of whether the perceptual content of that experience is veridical or illusory in a deep, metaphysical sense. Thus, it will suffice to presume that *something* in (or of) our subjective experience deserves to be called a self. (Herein, to keep the *psychological* and the *metaphysical* types of selves apart, I will stylize "psychological self" as SELF, and any general, non-specific type will remain un-stylized.) For my purposes, we needn't furnish a precise definition of this SELF; it serves here only as an umbrella concept, and subsumes the more specific types of SELVES on which this section and the next focus.

Thus, we can think of the SELF—again, in relation to memory—roughly as: (1) containing or comprising, in some way, some set of memories; (2) (putatively) unique to each individual, in a way that metaphysical selves are not, in virtue of how it relates to that set of memories;[30] and (3) perhaps most intuitively captured by the folk conception of the "soul" or "spirit" or "personality"—that which, besides your physical body or any other external characteristic, makes you *you*.

3.2.1 The Narrative Self

One common subtype of the SELF is the *Narrative Self*. Most ardently defended by Marya Schechtman,[31] the Narrative Self varies in its fine details among its proponents (Schechtman, 1996, 2011). Nevertheless, common threads run through all Narrative Self–conceptions: namely, that (1) the

[30] This is for the most part true, and is certainly true of most psychological selves we would expect to encounter in the world—although, as we will see in §3.2.2 (Rowlands's "Norman" example), it is possible, at least in theory, that two psychological selves of a certain type (viz., autobiographical selves) can be somewhat non-unique.

[31] Among others, the philosophers Alasdair MacIntyre (1984), Charles Taylor (1989), Paul Ricoeur (1994), Daniel Dennett (1992), David Velleman (2006), and Daniel Hutto (2008), as well as the psychologist Katherine Nelson (2003), also espouse multifarious narrative conceptions of the self.

SELF persists through a lifetime, each epoch of which relates to both each other and the whole as do chapters in a narrative; and (2) the SELF in one epoch relates to itself in all other epochs as does the protagonist in one chapter to himself in another. The Narrative Self, then, accounts for why, although one's *idea* of one's SELF can change drastically throughout a lifetime, one is not, at any later point, an entirely different SELF than at any earlier one. For one can, after all, "play back" one's narrative and trace the origin of one's current SELF-image through the changes—often minor and gradual, but sometimes abrupt—that brought it to this point. Borrowing Rowlands's wording, we might define the Narrative Self thus:

> **The Narrative Self:** A subtype of the psychological self "spliced together from the stories we tell about ourselves," whereby "the persons we … become have the same structure as a narrative." (2017, p. 86)

Immanent in the very concept of the Narrative Self (herein: NS) is its tight connection to memory; that a severe amnesiac could have an NS, on this definition, simply defies common sense.[32] Might the NS, then, be the target of our present inquiry? That is, since memory is so crucial to the NS, wouldn't precluding the consolidation of new memories via OMM be inimical to this SELF in some way? I don't think so, for two reasons.

First, the NS does not comprise the *contents* of our memories *per se*, but rather describes how our minds understand, evaluate, and organize these contents. For example, Habermas and colleagues glean from empirical studies on *narrative memory* that, as we mature, we develop the

[32] One might propose that she could, say, every morning, read a journal that she has kept throughout her entire life, one which describes every single event, sensation, emotion, mood, thought, hunger pang, daydream, and so on, in exquisite detail. But this seems more like perpetually *learning anew* the facts of one's life rather than *remembering* oneself narratively. Just as, every night, when I pick up the novel on my bedside table, I needn't read the entire book to the point where I last stopped in order to understand what is going on, I, as a Narrative Self, needn't continually review the entirety of my life to orient myself within its current chapter (although periodic reviews certainly cannot hurt). Some *extended mind* theories might be able to save the Narrative Self of a hypothetical amnesic patient in a similar way, although these theories remain contentious, and exploring them would take us far afield of this project's scope.

capacity to organize our memories into coherent sequences (Habermas et al., 2010; Habermas & Bluck, 2000; Habermas & Köber, 2015).[33] By adolescence, "individuals are able to *create* personally meaningful links among their experiences," note Fivush and Waters:

> These links include *creating* causal connections between events (e.g., "Because I went to computer camp, I became really interested in building robots, and that led me to join the computer club at school and build a robot for the science fair"), as well as autobiographical connections between events and self-understanding (e.g., "Winning the science fair was really important to me because it made me realize that if I work hard, I can achieve my goals"). (2019, p. 62; my emphasis)

Notice how they use the active-agential verb "creating" here. The NS is something we *do* rather than what, or who, we *are*; it's not so much an entity but an action—or at the very least, a *disposition* to act—and as such, raises questions: What, exactly, is doing the acting? What, exactly, is so disposed? These questions have yet to be satisfactorily addressed by narrative theory in general.[34]

Second, the NS is highly resilient to the loss of individual memories, whether lost naturally (viz., forgotten) or artificially.[35] Again, most people don't remember *most* of their lives, and yet

---

[33] The authors sometimes call this *autobiographical* memory. I will refrain from doing so, however, so as not to create confusion between narrative memory and another kind of psychological memory, which, as we will see momentarily, could easily (and perhaps more appropriately) take the title of "autobiographical" memory.

[34] Although, many narrative theorists—not all, but many—remain metaphysically agnostic in their analyses of the Narrative Self, just as I do here. What often seems to matter to them, as matters to me, is simply describing the phenomenological or cognitive experience of the self and, less often but not infrequently, connecting it to neurobiological or cognitive-scientific evidence, thus leaving the metaphysics to the metaphysicians, or (if they're not one and the same) to those unconvinced that phenomenology can, by itself, sufficiently justify talk of selves as such.

[35] The difference between "natural" and "artificial" forgetting is meant only to be intuitive at this point; I don't wish to get into the weeds over what, strictly speaking, counts as "natural" memory loss and what does not. But for now, we might intuit that a natural forgetting is akin to what Rowlands calls *passive forgetting*, or forgetting over time and without conscious awareness; we might then liken "artificial" forgetting to *motivated forgetting*, whereby one takes some sort of action to eliminate a memory from one's mind (2017, pp. 98-9). I would include in this latter category taking any sort of action to stop a memory from consolidating in the first place; thus, OMM would count as a means of motivated, non-natural, "forgetting."

can assemble coherent narratives from what relatively little they remember, or at least can be *prodded* to remember (more on this in §5). This point is most conspicuously revealed by the fact that people who have suffered severe memory loss have, indeed, written memoirs—have constructed *literal narratives*—about losing their memories.[36] Thus, unless one were to lose entirely her store of and capacity for memory, she could still forge a coherent narrative from the memories she does possess, even if that narrative is *about* (or partially about) losing her memory. In fact, it doesn't seem impossible (though perhaps scarcely feasible) that someone who's lost her mnemonic faculties could perpetually relearn her life story and understand it as a narrative—that, say, every morning, she could reconstitute her NS, like juice that has settled overnight.[37] But we're not concerned here with the (highly improbable) scenario wherein OMM causes global amnesia. We're concerned, rather, with precluding the consolidation of only brief episodes. So even if amnesia is in some way inimical to the NS, it is not inimical in the same way that, we should worry, OMM would be.

So, where has all this left us? Even if we grant OMM precludes some memories from consolidating, whether it harms the self remains unclear; for, as we've seen, OMM does not interact with the metaphysical self, and if OMM affects the NS at all, it does so only by affecting a *capacity* of some psychological self (SELF) rather than a SELF as such. One might note that affecting someone's capacities in this way could pose a kind of indirect harm to the SELF. But we needn't worry

---

[36] *See*, for instance: *On the Sea of Memory: A Journey from Forgetting to Remembering* by Jonathan Cott (2005), a former *Rolling Stone* writer who cannot account for anything he experienced nor any knowledge he gained in the fifteen years between 1985 and 2000, the result of neurological damage caused by extensive electro-convulsive therapy (ECT) for treatment-resistant depression (TRD; although this occurred in the late 1990s, when ECT was much less safe than it is today). Another account of ECT-related memory loss, also from the 1990s, is the psychotherapist Martha Manning's memoir, *Undercurrents: A Life Beneath the Surface* (1996), which details her struggle with TRD, hospitalization, and eventual ECT treatments. Manning is deeply funny, to boot—and a wonderful stylist.

[37] This seems theoretically possible only if, and to the extent that, different narrative theories grant both: (1) that a Narrative Self can be constituted entirely by semantic memories, but not at all by episodic memories; and (2) that the concept of "memory" can properly include facts, accounts of episodes, etc., that are not in some way stored or encoded into (and therefore decoded from) the matter inside one's head, as some extended mind theorists might contend, although I doubt that many outside that camp would accept this possibility.

about that yet. For now, we are still in search of a SELF that is *directly* and *proximally* affected by OMM, which is of greater concern than finding an *indirect* effect. The former should therefore remain our analytical priority, and we should only consider the latter if we have exhausted our present inquiry. Indeed, there is another subtype of SELF, which, although less popular than the NS, nevertheless deserves our consideration.

3.2.2 The Autobiographical Self

If the NS comprises an action (or a disposition to act)—a way of organizing (or of being disposed to organize) the contents of one's memory into coherent wholes—then what of the contents themselves? Might there be a conception of SELF that consists not in the *arrangement* of one's memories, the story one understands oneself to be living, but in those memories *simpliciter*—in the plain, unvarnished facts of one's life? Alas, such a conception exists. We can call this conception the *Autobiographical Self*, and define it thus:

> **The Autobiographical Self:** A subtype of SELF comprising the entire set of one's individual, subjective, declarative memories—both *semantic* and *episodic*, e.g., the semantic memory that one was born in Indianapolis, Indiana; or the episodic memory of a fishing trip with one's grandfather.

Let's illustrate the difference between the NS and the Autobiographical Self (herein: AS) with an analogy. Imagine that you have, spread out before you, a series of still photographs taken during different events in our lives. Of course, you can arrange the pictures in any order you'd like, so some of them you might cluster together; others you might arrange sloping upward, perhaps to represent a climax in your life; still others you might arrange sloping downward, to represent a nadir. If the *arrangement* of the pictures constitutes the NS, then the pictures *themselves* constitute the AS. The latter comprises the experiences you have had and the facts you know, the

pieces of the puzzle to which you're constricted in constructing your narrative. The final arrangement, however, can be multiple and varied. And yet, the AS is still not the kind of SELF so inextricably linked to memory that OMM poses a harm to it. Why not? Because it may not even constitute a SELF at all.

Whatever other ontological commitments we hold regarding the SELF, any proper conception should meet at least two criteria: It should explain, first, how that SELF remains the same entity through time, and second, the characteristics that distinguish it from other SELVES. We can call these the *Diachronicity Criterion* and the *Uniqueness Criterion*, respectively, and it is the latter that the AS cannot meet. To see why, consider the following example, again from Rowlands:

> You are writing a letter to a friend—let's call him Norman. You have, however, more than one friend named Norman. What makes the letter a letter addressed to the one Norman rather than the other? The letter, of course, has content. You describe a fishing trip you both went on last summer. But, as a matter of fact, you also went on a (different) fishing trip with the other Norman last summer. You add further content—describing the enormous trout that he, Norman, managed to catch. But, coincidentally, the other Norman also caught an enormous trout. You, presumably, see where this is going? You keep adding detail—content—to the letter, but that detail is equally appropriate to both Normans. This is certainly unlikely, but it does not seem impossible. In virtue of what, then, is the letter addressed to the one Norman rather than the other? The point is not, of course, that no possible detail or content could be sufficient to determine to whom the letter is addressed. Rather, the point is that it is an utterly contingent matter whether this content is included in the letter. (2017, p. 39; citing Malcolm, 1971)

Put another way: If, as the Autobiographical conception assumes, the SELF consists solely in the *Intentional contents* of one's memory[38]—in the events or facts which the memories are *about*—then as long as two or more individuals have either (1) experienced the same events and learned

[38] This stylization—with the capital "I"—is meant to differentiate the more colloquial concept of "intentionality," meaning, roughly, the quality of *being purposive*, from the philosophical concept of "Intentionality," meaning, roughly, the quality of *being about* or *referring to* something.

the same facts, or (2) each *forgotten* only those events and facts which the other has also forgotten, and *remembered* only those events and facts which the other has also remembered, they would have (or *be*) numerically identical SELVES, despite occupying two different bodies. Put yet another way: It follows from the principle of the Identity of Indiscernibles[39] that two such autobiographical selves—those "belonging to" two different human persons, in this case both named Norman—must, in fact, be one and the same AS. This possibility, however, is incompatible with the SELF, both intuitively and as we've defined it.

One might object that the AS *can* meet the uniqueness condition, because all episodic memories represent past events from a unique vantage point, one from which only a single individual could have experienced them. In other words, you cannot write a letter whose content can discriminate between Normans *not* because their respective sets of memories are identical, but rather, because you, the letter-writer, who also occupies a unique phenomenal perspective, cannot access the content that would pick out the correct Norman: namely, the experiential content that is *indexed to* (from the first-person perspective of) the correct Norman. But while this point about the indexicality of episodic memory may be true for *some* episodic memories, it is not definitionally true of episodic memory, nor is it likely to be true for all of any one person's episodic memories. It is not inconceivable that both Normans could, say, forget most or all of the episodes that constitute their autobiographical selves, leaving only the *semantic*—that is, *non*-indexical—memories intact. Thus, "whether the specific memories that distinguish the two Normans are retained or forgotten is an utterly contingent matter," notes Rowlands. "It is unlikely, but perfectly conceivable that the

[39] I.e.: If, for every property F, object x has F if and only if object y has F, then x is identical to y. Or, formally: $\forall F(Fx \leftrightarrow Fy) \rightarrow x = y$ (Forrest, 2016).

very . . . memories that would distinguish one Norman from the other have been forgotten—by both Normans" (2017, p. 40).

Nor is it the case that episodic memory, as a kind, must contain experiential content indexed to a first-person perspective. As Rowlands again points out, the phenomenon of *perspective switching* has been well documented, and undermines the view that episodic memory is always represented as its subject experienced it. Perspective switching occurs when one comes to remember some event that they experienced not from the perspective that they themselves observed it—that is, from their own perspective—but from the apparent perspective of some disembodied third-person; one can "shift from remembering an episode in which one was involved to remembering how one might have appeared to others as that episode unfolded" (Rowlands, 2017, p. 46; citing Goldie, 2012) . While, again, it may be unlikely, it is not inconceivable that the two Normans, even if their ASs still fully or partially comprise episodic memories, might still not possess (or *be*) unique ASs. If, for example, all of both Normans' episodic memories have taken on a third-person perspective, then you, the letter-writer, could no longer ascertain that content which would pick out the correct Norman, even if you *could* somehow access their episodic memories. Thus, the objection does not foreclose the possibility that two persons might possess identical sets of autobiographical memories, and the AS still cannot meet the Uniqueness Criterion.[40]

What might we add to the AS, then, to capture the conception of SELF we seek—the conception which is constituted, largely if not wholly, by memory? On its own, the AS comprises a mere collection of memories, which, as I've argued, anyone else's AS can theoretically comprise, and which can be ordered by the NS into multifarious arrangements. But none of that seems conceptually unique to any one SELF. How did *these* photographs come to sit before me? Whence

[40] "Identical" here as in *type* identical, which is all two sets of autobiographical memories need to be to fall short of meeting the Uniqueness Criterion.

these memories, in particular? The answer to these questions, which I take up in the next section, will cast into sharp relief the SELF we are after: that conception which best satisfies the Coherence and Uniqueness Criteria, and to which OMM is inimical.

4 THE PROUSTIAN SELF

> *Lulled in the countless chambers of the brain, our thoughts are linked by many a hidden chain. Awake but one, and lo! what myriads arise!*
>
> – Alexander Pope[41]

In his 1677 tome *Ethics*, the Dutch philosopher Baruch Spinoza writes presciently of memory. "Memory," he begins,

> enables the mind to perceive things through their first causes, and is the same in all men. Hence we can clearly understand how it is that the mind from the thought of one thing at once turns to the thought of another thing which is not in any way like the first. . . . In this manner each person will turn from one thought to another according to the manner in which the habit of each has arranged the images of things in the body. The soldier, for instance, if he sees the footsteps of a horse in the sand, will immediately turn from the thought of a horse to the thought of a horseman, and so to the thought of war. The countryman, on the other hand, from the thought of a horse will turn to the thought of his plough, his field, etc.; and thus each person will turn from one thought to this or that thought, according to the manner in which he has been accustomed to connect and bind together the images of things in his mind. (2008)

Spinoza's idea here sounds rather like *associationism*: the notion, often attributed to the Scottish philosopher David Hume, that our perceptions of objects in the world (what Hume called "impressions"), are stored in one's memory as concepts ("ideas") (Mandelbaum, 2017). Each concept then becomes associated—linked—to varying degrees, with another concept in memory. If I see the

[41] Pope's stanza here is quoted in full in (though, curiously, left unattributed by) Dierhold (1898, p. 238); I've reformatted the lines from lyric to prose for readability's sake.

color red, for example, I might be inclined to think about firetrucks or roses, because those concepts are nearby—strongly linked—to the concept "red" within the vast network of associations that constitute my memory. I won't, by contrast, be inclined to think of a dachshund or a plastic bag or a glass of water, since those concepts are, for me, comparatively distant—weakly linked—to the concept "red" than are the former concepts. Following Rowlands (2017, p. 15), I will refer to the cognitive, conceptual level of associative memory as the *personal level*, on which we need only talk about psychological states and dispositions rather than neurobiology.[42]

Nevertheless, associationism—or at least something very similar—has also been vindicated by the latest evidence in the neuroscience of memory. As we've already seen, we can roughly understand individual memories, as they inhere in the brain, as networks of constantly communicating engrams (*see* §2.1). Recall that these engrams are highly malleable, their component connections perpetually changing in *strength*: the ability of individual neurons in the network to activate others (Rudy, 2018, p. 176). Again, following Rowlands (2017, p. 15), I will refer to this empirical, neurobiological level of associative memory as the *subpersonal level.* Setting aside issues of the deep, metaphysical relation between the subpersonal and personal levels, minimally,

---

[42] This is, of course, not to say that there is only philosophical and (with respect to Proust's case below) literary speculation. For instance, the psychologist Richard Holmes and his student, Marisse Clark, tested the effect of *memory boxes* on elderly patients in Norwich, United Kingdom, which elicited remarkable—and, if I may, rather heartwarming—results. Rowlands describes the study aptly:

> The women were asked to describe memories of their youth in the 1930s and 1940s. Initially, their memories were scattered, fragmented and it was difficult to elicit more than a few well-worn tales. This changed significantly with the introduction of the memory box: in this case, a large suitcase containing a number of mundane objects from the 1930s—a bar of Lux soap, an Ovaltine tin, a packet of Swan Vesta matches, a small mangle, a tortoise-shell hair clip, a stone hot water bottle and so on. Holmes comments: "The effect of the memory boxes was often magical. The old women, many in their seventies and eighties, slowly began to handle, identify and discuss these familiar objects. Amazement was soon followed by laughter, delight, and not infrequently indignation, and even some tears. Each physical object began to 'trigger' a long chain of recollections. Gradually, an extraordinary stream of shared memories, anecdotes, jokes, and stories would emerge. The flow—the floor—soon became unstoppable." (2017, p. 18, fn. 23; citing Holmes, 2008)

we may take the latter to *supervene* on the former. That is, we can take personal-level properties to result from subpersonal-level properties and any change in the former to entail a change to the latter—though not necessarily vice versa (McLaughlin & Bennett, 2018).

So, going from the subpersonal level to the personal, we might represent neural engrams as *nodes* within a network, which themselves correspond to some content in memory. Each engram's various synaptic connections with other engrams we might represent as connections, varying in weight (or strength or thickness), between nodes. Thus, on the subpersonal level, the engram which physically represents my concept "red" would have relatively strong synaptic connections with my engram physically representing the concept "firetruck." This would translate, on the personal level, to my "red" node being strongly linked with, or tending to elicit, my "firetruck" node. But note that the malleability of engrams means that one's associative concepts and the relations between them are frequently in flux. Although my "red" node currently tends to elicit my "firetruck" node, in the future, I might experience, say, an especially grisly injury, causing my "blood" node to become so strongly connected to my "red" node that now, the latter tends to elicit the former. This is known, subpersonally, as *reconsolidation* and *trace updating* (Rudy, 2018; esp. at pp. 275-277; Suárez et al., 2010; more on reconsolidation in §4.2).

The basic idea of trace updating is that, as one ventures through life, one constantly incorporates new experiences and knowledge into, and thereby modifies, the total set of engrams that constitutes one's total memory. Because engrams, subpersonally, correspond to memories, personally, it is this sprawling set of engrams that constitutes one's AS; for it contains everything one could at least theoretically recall in order to give a full, narrative account of one's life. But, as we've also already seen (Rowlands's "Normans" example), even one's total set of memories (engrams, subpersonally speaking) cannot *on its own* constitute a proper SELF, as it cannot satisfy the

Uniqueness Criterion for SELFhood. Rather, echoing Spinoza, it is the *connections* between these memories, between these engrams and their components, as well as the *manner* in which one's phenomenal, mnemonic experience is apt to unfold along these connections, which constitute the SELF we've been pursuing: the SELF so tightly intertwined with memory that any change to the latter will necessarily change the former, and thus, the SELF that is principally threatened by OMM. This, I will call the *Proustian self* (*see* Fig. 4-1).

I've named the Proustian self after the French author Marcel Proust, whose *magnum opus* is the multivolume work *À la recherche du temps perdu* (*In Search of Lost Time*), in which the unnamed narrator—usually called "Marcel," after his creator—guides readers through memories of his childhood and adolescence in France at the turn of the 20th century.[43] It will be useful at first to give a phenomenological description of the Proustian self—and although the work is replete with vivid descriptions of memory, I find the following passage well suited to this task.[44] Near the beginning of the first volume, Marcel, the narrator, recounts the fateful moment he was first thrown back into the memories of his youth:

> Many years had elapsed during which nothing of Combray . . . had any existence for me, when one day in winter, on my return home, my mother, seeing that I was cold, offered me some tea, a thing I did not ordinarily take. . . . She sent for one of those squat, plump little cakes called "petites madeleines." . . . I raised to my lips a spoonful of the tea in which I had soaked a morsel of the cake. No sooner had the

[43] At the time of this writing (March 2020), the philosophical literature contains only one other use of the term "Proustian self": Joshua Landy's piece in *New Literary History* titled "'Les Moi en Moi': The Proustian Self in Philosophical Perspective" (2001). In it, Landy performs a masterful literary analysis spanning all six volumes of *A la recherche du temps perdu*, primarily to draw out the notion of selfhood suggested by Proust throughout the work. The analysis is multilayered and complex, and is, in some ways, congruent to the Proustian self that I describe here. But it is also different in many crucial respects. However, because an exhaustive comparison between his and my conceptions would require an entire paper unto itself, I will not undertake such a project here. For now, I will only point readers toward Landy and encourage them to compare my and his "Proustian selves" on their own.

[44] I reproduce this passage fully aware that it is perhaps the most oft-reproduced bit of prose in human history—besides, perhaps, some Shakespearean soliloquy or Biblical allegory. My hope in doing so is not to perpetuate a cliché, but rather to remind readers of the almost singular phenomenological brilliance of this passage, which I could not hope to replicate through description alone.

> warm liquid mixed with the crumbs touched my palate than a shudder ran through me and I stopped, intent upon the extraordinary thing that was happening to me. An exquisite pleasure had invaded my senses, something isolated, detached, with no suggestion of its origin. . . . Whence could it have come to me, this all-powerful joy? I sensed that it was *connected* with the taste of the tea and the cake, but that it infinitely transcended those savours, could not, indeed, be of the same nature. . . . I decide to attempt to make it reappear. I retrace my thoughts to the moment at which I drank the first spoonful of tea. . . . I place in position before my mind's eye the still recent taste of that first mouthful, and I feel something start within me, something that leaves its resting-place and attempts to rise, something that has been embedded like an anchor at a great depth; . . . Undoubtedly what is thus palpitating in the depths of my being must be the image, the visual memory which, being *linked* to that taste, is trying to follow it into my conscious mind. . . . And suddenly the memory revealed itself. The taste was that of the little piece of madeleine [cake] which on Sunday mornings at Combray . . . when I went to say good morning to her in her bedroom, my aunt Léonie used to give me, dipping it first in her own cup of tea or tisane. The sight of the little madeleine had recalled nothing to my mind before I tasted it; . . . But when from a long-distant past nothing subsists, after the people are dead, after the things are broken and scattered, taste and smell alone, more fragile but more enduring, more unsubstantial, more persistent, more faithful, remain *poised* a long time, like souls, remembering, waiting, hoping, amid the ruins of all the rest; and bear unflinchingly, in the tiny and almost impalpable drop of their essence, *the vast structure of recollection*. (Proust, 1913, pp. 48–50; my emphasis)

Note the similarities between Spinoza's passage and Proust's. First, the subject perceives some stimulus: in Spinoza's case, the horse's footsteps in the sand; in Proust's, the *petite madeleine*. This perception then initiates a cascade of autonomously unfolding mental representations. Proust likens these unfoldings to the Japanese art form of dipping formless pieces of paper into a bowl of water, then watching them "stretch and twist and take on colour and distinctive shape, becom[ing] flowers or houses or people" (1913, p. 51). For Proust's "Marcel," these unfoldings ultimately take the shape of his childhood hometown, Combray, where *A la recherche du temps perdu* begins in earnest. Additionally, the unfolding mental representations are *linked* or *connected* in some way—though, for both Spinoza and Proust, it is not entirely clear how so. What *is* clear is that, at least

phenomenologically, one representation begets another, which begets another and another. It needn't be the case that what lies at the end of this chain is closely connected to what started it; for Spinoza's soldier does not jump directly from thoughts of horses to thoughts of war, nor does Marcel jump from the *madeleine* to Combray. Rather, each representation is closely linked to both its direct precursor and descendant: horse-thoughts beget horseman-thoughts, and *madeleine*-thoughts beget thoughts of aunt Léonie.[45]

4.1 Coming to terms

It will be useful at this point to lay out some stipulative definitions. Following Conway and Loveday (2010), we'll call each individual, *episodic* scene that unfolds in the mental theater of memory is an *episodic element* (EE). With respect to *semantic* memory, each fact, bit of information, proposition, meaning, and so forth, we'll call a *semantic element* (SE). When we needn't distinguish between the two (which will be most of the time) we can call either, simply, a *mnemonic element* (ME). For example, should the thought "these are the hoof-prints of an enemy horse" arise in the mind of Spinoza's soldier upon seeing the prints, it may not be the case that the notion of "enemy horse" consists of a singular, specific episode involving an enemy horse (since he has presumably encountered innumerable enemies' horses). In that case the thought would contain an SE, consisting solely of semantic content. Should the thought "plough" arise in the mind of Spinoza's countryman, by contrast, it may be the case that he does think of a specific episode involving a plough—perhaps he was once traumatically injured by one. In that case the thought would

[45] "Linked," that is, not *conceptually* or *metaphysically* (i.e., mind-independently), whatever that might mean, but in a manner that would make sense to the subject, or at least *could* make sense to the subject, given, perhaps, a jogging of memory, or a psychotherapeutic insight into some past experience, or some other such prodding.

contain an EE, consisting solely of episodic content. Both SEs and EEs, and, if possible, any combination of the two, fall under the broad category of MEs; thus, Marcel's vision of his aunt (EE) as well as his thought that his aunt's name is (or was) Léonie (SE) are both MEs.

A *cue* is a stimulus that elicits the unfolding cascade of MEs. Spinoza's hoof-prints are a cue, as is Proust's *madeleine*. Cues needn't be sensory or environmental, however—MEs can also be cues. Thus, the *madeleine* is a cue, which elicits the ME of his aunt Léonie; but that ME in turn elicits the ME of she and little Marcel dipping *madeleines* into their tea, and is therefore also a cue. Cues are thus more a *functional* than a substantive concept. Sometimes, either spontaneously or when one gains conscious access to what one was trying to remember, an ME will arise which elicits no further MEs. So, while MEs *can* be cues, they needn't necessarily be.

Linkages are the connections between MEs, the *strength* of which denote the likelihood that one particular ME will follow from another, or the disposition of one ME to cue another. These linkages *physically* inhere in the synaptic strength (the ability of one "node" to activate another, to reuse our circuit analogy) between neuronal connections within and between engrams. Linkages *psychologically* inhere in the proximity of associated concepts (as in "red" and "firetruck") specific to each individual. Figure 4-1 presents a visual of this notion: The soldier (A) has strong linkages between "horse," "horseman," and "war," for example, as well as between "frog" and "green," and "field" and "plough"; likewise, the countryman (B) has strong linkages between "horse," "field," and "plough," as well as between "frog" and "green," and "horseman" and "war," though not between "horse" and "horseman," or "field" and "green." Moreover, since strength here is a *dispositional property*—a *potentiality*—rather than a substantive one, these linkages needn't be actively engaged to be strong or weak.

From this initial impression and delineation of terms, let us roughly define the Proustian self as follows, after which we'll shade in its fine details:

> **The Proustian Self:** A subtype of the SELF, constituted by both (1) the total set of one's MEs, and (2) the strength of the linkages between those MEs—which manifests as the manner in which, at any given time, those MEs are disposed to arise by virtue of a *cue* (whether a sensory cue or another ME) and, in turn, cue other MEs, which can cue others, and so on, in an iterative fashion.

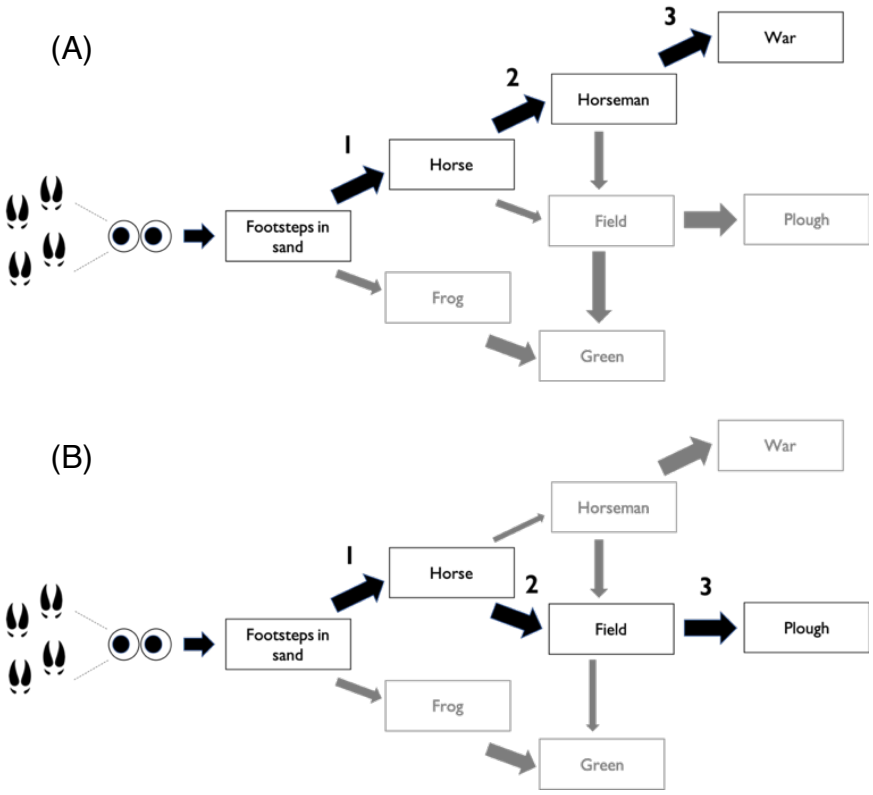

Figure 4-1 *A simplified representation of two Proustian selves. We might take (A) to represent Spinoza's soldier, (B) his countryman. When (A) sees horse-hoof prints in the sand (the* cue*), he experiences an "unfolding" of episodic or semantic mnemonic contents (*EEs *or* SEs*, respectively, both of which are, broadly,* MEs*), along a specific path (i.e., "horse"–"horseman"–"war"), which is determined by the strength of the connections, or* linkages*, between those* MEs*, which are themselves physically underwritten by specified neuronal networks, called* engrams*. Note that, on this conception of the* SELF*, while (A) and (B) might comprise numerically identical* MEs *(and are, therefore, numerically identical autobiographical selves, as in Rowlands's "Normans" example), they are nevertheless different Proustian selves, for each one experiences different memory unfoldings, even if cued by the same stimulus.*

Thus, on this conception of the self, what makes Spinoza's soldier *himself* is that he is disposed, first, to experience a certain type of "mnemonic unfolding" when he sees (or tastes, smells, hears, etc.) a cue, horse hoof-prints or otherwise, and second, to react a certain way in virtue of that unfolding; perhaps his memory reels in search of where he's seen that particular hoof-bring before, whether it was made by the horse of an adversary or an ally, so that he can either track it or leave it be. What makes the countryman *him*self is that he is otherwise so disposed. Neither of them, likewise, could nibble on a pastry and experience the buildings and streets of Combray erecting themselves in consciousness just as Marcel does, nor be similarly disposed to recount the same story, in the same arrangement and style—even if we stipulate that they, too, grew up there and that they, too, had aunts named Léonie with whom they ate *petites madeleines* dipped in tea.[46]

Now, recall that we're trying to discern how the SELF is constituted by memory, or, put another way, which conception of the SELF one needs one's full mnemonic capacities to sustain. So before showing how the Proustian self (herein: PS) is this SELF (and is therefore the SELF threatened by OMM), we need to show that it is a SELF *in the first place*. That is, we need to show that it meets both the Diachronicity Criterion and the Uniqueness Criterion. Let's take the latter criterion first.

4.2 The Uniqueness Criterion

The PS satisfies the Uniqueness Criterion precisely because it has what the AS lacks. Recall that the AS—and by extension the NS—does not fit the Uniqueness Criterion because it is conceptually

[46] Unless, of course, the soldier *is* the countryman, and is also Marcel.

possible for two autobiographical selves to comprise indiscernible mnemonic contents, as Rowlands's "Normans" example illustrates.[47]

This problem arises, fundamentally, from a mistaken conception of memory: namely, as Rowlands points out, "the [erroneous] identification of the content of a memory with *what* one remembers" (2017, p. 7, fn. 11; author's italics). This is because the Intentional contents of memory (the events or facts to which they refer) are *external* to the individual rememberer—nothing about the individual human being or anything that he or she does, on the mistaken picture, makes his or her memories, and thus the AS that comprises those memories, unique. And a SELF that is not unique simply doesn't square with any intuitive notion of the self, nor with the SELF as we've defined it. We might call this mistaken notion *mnemonic-content externalism*. The conception of the PS, by contrast, does not entail mnemonic-content externalism, for two reasons: (1) unlike the AS, the PS consists not only in the *contents* of memory but also in the associative *relations* between those memories, and (2) on this conception, the very *act of remembering*—circumscribed entirely by the individual's *internal*, subjective experience—is a crucial constituent both of mnemonic contents and their associative relations. We can unpack this idea on both the personal and subpersonal levels.

To unpack this on the personal level, we will again take our cues from Rowlands (2017). Prior to explicating his conception of the self, Rowlands makes numerous important distinctions on phenomenology and the nature of cognition, which we can also adduce to my argument. Rowlands's crucial point is that the "Anglophone philosophical world" has perverted the phenomenological method, stripped it down and denuded it of the analytical utility that its Continental

[47] Again, since, for my purposes, these selves needn't exist metaphysically, but only conceptually or phenomenologically, conceptual possibility is all we need.

founders (Husserl, Heidegger, Sartre, Merleau-Ponty, among others) understood it to have (2017, p. 21). Phenomenology, as it is widely understood today, erroneously takes as its objects the *contents* of perception (including introspection), or what "appears" in experience—what some philosophers call *qualia*. In turn, some contemporary philosophers of mind apply this misunderstanding of phenomenology to consciousness *itself*, and thus try to demarcate consciousness by virtue of how it appears to perception. In other words, they take phenomenology's object to be the Intentional objects of perceptual experience, and then simply define consciousness as "what-it-is-like" to perceive those objects.[48]

But, as Rowlands argues, this is not the *end* of the method. "[F]rom the perspective of the [original] phenomenologist," this contemporary notion is a metaphysical false bottom, as it were, and therefore "any view of conscious experience grounded in it is incomplete" (2017, p. 24). On its original conception, the phenomenological method's starting point (known as the *epoché*, or *phenomenological reduction*) indeed takes as its object the Intentional objects of—the appearances within—experience, but then acknowledges that "appearances presuppose something to which these appearances appear," something which, logically, cannot itself show up in experience (Rowlands, 2017, p. 25).[49] The move, then, is to say that that *of* which we are aware is different from that *with* which we are aware; and from there, it is a short step to the idea that the object of phenomenology is the former *in virtue of* the latter. Rowlands puts the point aptly:

> Phenomenology is the discipline that is concerned with studying mental acts precisely as mental acts—rather than as objects of mental acts. . . . We begin with appearances, and then we try to work out how mental acts must be if appearances are to be the way they are. (2017, p. 27).

[48] This kind of view Rowlands attributes to philosophers such as Colin McGinn, Frank Jackson, and Thomas Nagel, surely among others.
[49] Cf. Duncan (2019).

And therefore, with regard to a proper phenomenology of memory, "understanding how 'our memories make us who we are' requires refocusing on the importance of the act of remembering and the role this plays in the *construction of the memory's content*" (2017, p. 27; my emphasis).

It is in light of this more robust (and, I believe, correct) conception of phenomenology that the uniqueness of the PS comes into view. We can see now that the conception of the AS entails the mistaken "Anglophone" notion of phenomenology. This is because the AS is defined solely as the total contents of—the Intentional objects of, or the appearances within—one's memory. Recall the Norman example. The letter writer is attempting to pick out one of the Normans by referencing events or objects which might appear in one or both Normans' conscious memories, but *not* anything putatively internal or subjective to either. But, just like any other mental act, one's memory, understood phenomenologically, must comprise not merely the contents of experience but the *way in which* one experiences those contents.

This conception of memory (which, I believe, is the correct personal-level conception of memory, because supported by the correct conception of phenomenology) adduces to the PS over competing conceptions, such as the AS. While the other conceptions mistakenly entail that the SELF (insofar as it is constituted by memory) comprises one's total mnemonic contents, which only *contingently* attach to a particular person's memory, the PS entails, correctly, that the SELF comprise both one's mnemonic contents and how those contents are constructed by the subjective mental *act of remembering* them, which necessarily attaches to a particular agent. In other words, the PS comprises one's mnemonic contents *qua* artifacts of one's synchronic subjective experience. Thus, insofar as subjective experience is unique to one person—indexed to one human being, who cannot occupy the same space at the same time as another—one's PS is unique to one person. And since the former clause is trivially true, the PS necessarily satisfies the Uniqueness Criterion.

The notion that synchronic experience continually interacts with memory is evidenced, as well, at the subpersonal level—namely, by *reconsolidation theory*. Recall that, theoretically, it is the reactivation of cortical engrams that issues in a conscious memory. One might imagine (and indeed, it is commonly believed) that one's memories are relatively stable, that the memory has, as it were, carved grooves into our brains, by which we can, like a record, replay them over and again. But this is not so. In 2003, the neuroscientist Karim Nader and his team demonstrated two phenomena: (1) "that when a memory is retrieved *the synapses underlying the [engram] become unbound or weakened*," meaning that the act of remembering "itself can disrupt an established memory trace," and (2) that this act of remembering, additionally, initiates the process whereby engrams are formed ("protein synthesis") and thereupon, "*the [engram] is reconsolidated*" (Rudy, 2018, p. 269; author's italics. *See*: Debiec & Nader, 2004; Nader, 2003; Nader et al., 2000). Put simply, the act of remembering some event X makes the memory of that event *labile*; it brings its engram back to the weakened state in which it was first constituted. But before it is reconsolidated, one's occurrent experience is, so to speak, folded into that original engram, and thereby changes it, however slightly. In this light, memories needn't be initially forming to be modified. For example, researchers have suggested that, perhaps, the environmental stimuli which cue drug addicts to use can be recalled into memory, their engrams then pharmaceutically (or otherwise) weakened, and thus deprived of their ability to elicit drug-seeking behaviors in the future (Lee et al., 2005, 2006).[50] Thus, reconsolidation theory not only accords with the personal-level notion that, phenomenologically, the experience of memory necessitates that only one PS necessarily attaches to

[50] *See* also: Dudai & Eisenberg, 2004; Finnie & Nader, 2012; McKenzie & Eichenbaum, 2011; Przybyslawski et al., 1999; and Tronson et al., 2006.

each individual person but also seems, subpersonally, to *evidence* it, as it holds that the experience of remembering the past always enfolds one's (unique) experience of the present.

4.3 The Diachronicity Criterion

The Diachronicity Criterion for SELFhood holds, roughly, that a SELF must persist through time, and thus requires that any conception of the SELF spell out in what that persistence consists. Perhaps the most thoroughgoing treatment of the issue of identity through time is Derek Parfit's (1984) discussion of the criteria of personal identity, so I will take it as instructive. Now, I should note that although Parfit discusses *personal* identity—and thus, insofar as his conception of persons differs from my conception of SELVES, his conceptual target is different than mine—his criteria for identity through time (viz., diachronicity) can sufficiently *generalize* to mine, especially since his conception of persons, like my conception of SELVES, turns mainly on how we understand the relations between our memories as psychological and physical states. Again, I think we can establish diachronicity on both the personal and subpersonal levels, but, in the interest of space, we will restrict ourselves to a personal-level explication here—of course, keeping in mind the supervenience relation we're taking to hold between conscious memory and its neurological underpinnings. My argument for why the PS satisfies the Diachronicity Criterion is mostly congruent with Parfit's explanation of personal identity through time, but turns on one crucial distinction. Thus, it will be useful to carefully enumerate his claims, but I will be sure to flag where, from Parfit, we will take our departure.

As Parfit points out, "it is memory that makes most of us aware of our own continued existence over time" (1984, p. 205). This is strikingly exemplified by Clive Wearing, whose inability to make new memories led him to believe he was perpetually waking from a deep, unconscious sleep. One might thus be tempted to think, as John Locke thought, that one persists from

time t_1 to time t_2 only if, at t_2, one remembers t_1 (Locke, 1979; Chapter 27, §16). On Locke's view, I am the same person who, say, nearly drowned in a boating accident when I was five years old only if I *remember* that experience. In other words, for one to persist from time t_1 to a later time, t_2, then at t_2, one must remember whatever happened at t_1.

But Locke's view cannot be true. I do not specifically remember writing the introduction to this paper, but it would be preposterous to assert that I therefore didn't write it. A modified version of Locke's view holds that there needn't be any *direct* mnemonic connections between me at t_2 and me at t_1 for me to persist from the latter to the former, but there must be *overlapping chains* of such connections. For example, if today I remember eating a sandwich yesterday, and yesterday I remembered writing the introduction to this paper the day before, then I have an overlapping chain of direct memories from today to three days ago, when, we'll imagine, I wrote the introduction. Parfit calls this *continuity of memory* and explains that "[o]n the revised version of Locke's view, some present person X is the same as some past person Y if there is between them [psychological] continuity of memory" (1984, p. 205).

We must now delineate two general relations: (1) *psychological connectedness*, which Parfit defines as "the holding of particular direct psychological connections," and which, for our purposes, we can understand as *mnemonic connections*, or diachronic connections between memories;[51] and (2) *psychological continuity*, defined as "the holding of overlapping chains of *strong (psychological) connectedness*" (1984, p. 206). So, generalizing from Parfit's view, we may now define the *diachronicity* of SELVES thus: Some present SELF X is the same past SELF Y if there are between them strong overlapping chains of particular direct mnemonic connections. Because

[51] For Parfit, psychological connections over time can also include connections between desires, beliefs, attitudes, motivations, etc. Since we're only concerned with how memories are connected over time, however, we needn't consider these other psychological entities, since we're concerned only with memory.

strength here is a matter of degree, I will follow Parfit in claiming, conservatively, that "there is enough connectedness [to count as *strong*] if the number of direct connections, over any day, is *at least half* the number that hold, over every day, in the lives of nearly every actual person" (1984, p. 206; author's italics). Some present SELF X is thus the same past SELF Y if, over every day, there remain between them at least half the number of particular direct mnemonic connections;[52] and thus, with respect to the Proustian self, my present PS (X) is the same as my PS yesterday (Y) if X comprises at least half of the mnemonic elements (MEs) that Y comprises, and Y is the same as my PS the day before (Z) if Y comprises at least half of the MEs that Z comprises, and so on.

The final important notion for coherence concerns the *causation* of memories. Parfit consolidates the foregoing considerations into what he calls the *Psychological Criterion*:

> (1) There is *psychological continuity* if and only if there are overlapping chains of strong connectedness. X today is one and the same person [*mutatis mutandis* for the SELF] as Y if and only if (2) X is psychologically continuous with Y, [and] (3) this continuity has the right kind of cause. . . . [(4) Diachronicity] just consists in the holding of facts like (2) to [(3)].[53] (1984, p. 207)

For Parfit, and I think intuitively, "the right kind of cause" for psychological (mnemonic) continuity would be any cause that is not "produced by abnormal interference, such as direct tampering with the brain" (1984, p. 207). On this view, if my present memory of playing with my dogs yesterday was caused by my playing with my dogs yesterday, and not, for example, by a mischievous

---

[52] Thus, subpersonally speaking, strong connectedness consists in the maintenance of at least half the number of accessible engrams (viz., those able to be activated, or not, for whatever reason, foreclosed from activating) "that hold over every day, in the lives of nearly every actual person." Although, of course, many, many more than half the number of such engrams (more specifically, their constituent strengthened synapses, given reconsolidation theory [*see* §4.2]) will hold from day to day in nearly every person, lest they've experienced some egregious neural insult in the intervening days or have some disorder whereby engrams cannot remain stable from one specified time period to the next.

[53] For my purposes, I've elided Parfit's fourth sub-criterion: "(4) it has not taken a 'branching' form" (1984, p. 207). Branching has to do with Parfit's previous discussions on the metaphysics of personal identity and time, and thus doesn't concern our discussion of SELVES.

neuroscientist implanting that memory, then it has the right kind of cause. All of the foregoing I take to be highly plausible conditions of SELF-diachronicity.

But now, we must depart from Parfit, and here's why. We've already seen how the PS works: internal or external cues set off constellations of unfolding MEs, which act as cues for further MEs—this unfolding being determined by the strength of the associative linkages between the MEs, which are themselves underwritten by past experience and modified by subjective experience at the time of recall (as we saw in the previous subsection). Thus, the PS fits the Diachronicity Criterion (*qua* Parfit's Psychological Criterion of personal identity through time, *mutatis mutandis* for the SELF) precisely because, by virtue of its mechanism, it *guarantees* overlapping strong connectedness and continuity (conditions 1 and 2). In other words, because the PS consists in extant MEs eliciting further MEs virtually moment by moment, it *necessarily* maintains strong connectedness, and does so to a degree well over Parfit's fifty-percent threshold. The PS also meets the Diachronicity Criterion because this continuity and connectedness is caused in the right kind of way; for what else but its inherent causal mechanism could count as the "right kind of cause" of this connectedness and continuity?

Interestingly, Parfit does *not* think that personal identity consists in the Psychological Criterion. Why not? Because he thinks that psychological continuity, or something "as good as" psychological continuity, can obtain even if caused by *non-natural* means, such as neural manipulation (Parfit, 1984, p. 209). He gives the example of artificial vision, maintaining that one's visual experiences could result from light waves hitting, say, a robotic eye connected to a neural implant, and that the psychological states resulting from such a process—for example, the memory of "seeing" a flower—could still provide the person continuity, despite being caused non-naturally. But this objection doesn't generalize to the Diachronicity Criterion. In the case of the

artificial eye, it is the *vision* which is unnaturally caused, not the *memory of the flower*—not, that is, the ME. If the existence of an ME *itself* were caused by neural manipulation—say, an artificially implanted ME of me playing with my dogs—then *that* would plausibly count as an ME that was not caused in the right kind of way. And any continuity requiring *this* kind of unnaturally caused ME (one manufactured out of whole cloth) cannot reasonably constitute SELF-continuity. So, to follow Parfit as far as he goes here and require that, to meet the Diachronicity Criterion, not only must a SELF have strongly continuous MEs, but its MEs cannot be constituted, even in part, by sensory prosthetics (such as artificial eyes) or some other sensory-mediation mechanism, would render the Diachronicity Criterion far too stringent. Thus, the Diachronicity Criterion should, at least for SELFhood, remain as originally defined.

In sum: (1) the Diachronicity Criterion (for SELFhood) is analogous to the Psychological Criterion (for personal identity); (2) the Diachronicity Criterion is not, however, vulnerable to Parfit's "right causes" objection, and thus, whereas the Psychological Criterion is not (at least for Partfit) an acceptable criterion for diachronic personal identity, the Diachronicity Criterion *is* an acceptable criterion of diachronic SELF-identity; (3) the PS, in virtue of its very constitution, meets Psychological Criterion; (4) therefore, the PS meets the Diachronicity Criterion.

5 TOWARD AN ETHICS OF FORGETTING

> *The past is hidden somewhere outside the realm, beyond the reach of intellect in some material object (in the sensation which that material object will give us) which we do not suspect. And as for that object, it depends upon chance whether we come upon it or not before we ourselves must die.*
>
> – Marcel Proust[54]

[54] 1913, p. 48.

In this section, I will bring us back to where we began, the ethical quandaries we first posed about the prospect of neuromodification now in sharper view. My aim here is not to give an ethical analysis of therapeutic OMM, as that would require a paper unto itself. Rather, I wish to demonstrate how we might begin to assess the morality of modifying memory, given its potential knock-on effects on the SELF, among any others. In other words, my aim here is simply to prepare the ground-work for the ethical analyses to come on this prospect as it inevitably approaches fruition.

5.1 A lockbox in the mind

The psychologists Martin Conway and Catherine Loveday (2010) tell the story of one of their patients, CR, who, in her mid-40s, contracted the same disease that ravaged Clive Wearing's memory, although its damage was less severe in her case. Though CR's medial temporal lobe, the area which houses her hippocampus, was largely destroyed, she nevertheless scored highly on tests of intelligence, executive function, and even short-term memory showed normal function. Past about half-an-hour, however, her memory began to fade—past three days, it was gone. In their article, Conway and Loveday quote CR as she feebly attempts to recollect, unable to recount her son's sixth birthday, only six days prior:

> Oh well, I can't remember very much about it. I think he was around but not awake at lunchtime—I don't think he got up until later on in the afternoon. But I really can't remember now.[55] (Conway & Loveday, 2010, p. 64)

Here, CR attempts to recount a night with her in-laws a week prior:

> Yes—I think I had a nice evening with them—yes. But what specifically happened—if anything—I'm not sure. . . . [P]erhaps I more likely helped L in the kitchen. If I went for the evening I think perhaps we'd have sorted tea so it wasn't a teatime evening—I don't know perhaps we all had a glass of wine—I'm not sure. I can't be sure what happened at all. (Conway & Loveday, 2010, p. 63)

[55] I've removed items like "erm," "um," and "oh," as well as other verbal tics like self-corrections, stutters, and pauses which Conway and Loveday indicated with ellipses. Where I've redacted actual words and phrases, I use my own ellipses.

Curiously, though, her family describes her memories not as *missing*, but rather, as "locked in a vault" (Conway & Loveday, 2010, p. 64). And indeed, when the researchers presented her with a *cue*, some object related to the event in question, her memories came back in remarkably vivid detail; she would proceed to recall with alacrity exactly what she did, to whom she spoke, what she bought, and so forth:

> So I picked up B and E and went up to the house. We walked up—I think. Yes I think we did [because] I didn't need the car. No—we didn't—we drove up thinking that if it took a long time helping unload and what have you then I would at least have the car there because I had to fetch the children from the school bus stop later. . . . And one of the things that I had done with I went to get the keys in Market Deeping—which I forgot to say just now—was I had walked back through past the florist and bought a nice bunch of flowers—some roses and other little pink and purple freesias. And I'd also spoken to the lady there about some flowers for the Golden Wedding Anniversaries coming up next week before I left. … And the other thing I bought when I was in Market Deeping in a little shop was a nice "new home" card. (Conway & Loveday, 2010, p. 64)

CR's case (among others; *see* note 41) suggests that our memories are rather unlike the file drawers or hard drives to which antiquated metaphors have likened them. Were it like these, a file or two could be destroyed—perhaps even entire folders—with little consequence for the integrity of the collection as a whole. No, our memories, and by extension, much or the whole of our SELVES, are more like vast libraries, complex labyrinths of information readily accessible only by reference to a much simpler index. Indeed, the *indexing theory of hippocampal memory*, a widely supported theory of memory storage and retrieval, makes perfect sense of CR's case, and in turn, the harm of OMM (*see*: Rudy, 2018, Chapter 16).

Consider how libraries solve the problem of storing lots of information storage while making retrieval as efficient as possible. When you go into a library, you are looking for some particular

information, your likelihood of finding which in short order is exceedingly small. Instead of ambling through the stacks, then, you begin by searching the indexing system, which tells you where to find the information you seek. Crucially, on this picture, none of the information, or *content*, you seek exists within the index itself, but in the book to where the index directs you. "Likewise," writes Rudy (2018), "the episodic memory indexing theory assumes the rich content of our experience is stored in neocortical regions of the brain and all that the hippocampus stores is information about how to retrieve the memories stored in the neocortex" (pp. 310-311). In addition to the phenomenological evidence of our occurrent experiences folding themselves into our memories (*see*: §4.2 on reconsolidation), we know that, at the subpersonal level, our engrams consistently incorporate new information from our occurrent sensory experience (Rudy, 2018, pp. 268–271). Therefore, since we also know that new information consolidates into long-term memory first by way of the hippocampus, we might surmise that certain manipulations or disorders that affect the hippocampus affect this *indexing system* rather than the storage of any *mnemonic content*.

In CR's case, for instance, we might theorize that the particular nature of her neural trauma left some long-term memory systems intact while damaging others, such as the hippocampus (index). Thus, her memories do not simply erase themselves over time, but instead become disconnected to the cues which would otherwise occasion them, and thus, remain locked away, unable to recall without first reencountering the cue—as this section's epigraph so gracefully describes.

5.2 OMM and the Proustian Self

Consider how memory, OMM, and the SELF interact. We know that MEs move from short-term to long-term storage by virtue of hippocampal consolidation. But this happens largely unconsciously;

while there are ways raise the likelihood that a long-term memory will form—for example, effortfully focusing one's attention on the fine details of an object, or telling oneself a story about it—we do not necessarily know which MEs our brains will find salient enough to consolidate into our network of MEs, and to become integral parts of our selves. Moreover, we do not know which aspects of our occurrent experiences will, at any given time, become *cues* for the recall of further MEs stored in the neocortex or parts of those neocortical engrams themselves. The likely effect of blocking memory consolidation, then, is that we absolutely foreclose the ability of any ME that might have consolidated during the period of modification from doing just that. Of course, this is the *point* of such therapy, at least in our anorexia example. The problem, however, is that we do not know—we *cannot* know—why our brains might have designated certain experiences appropriate for consolidation, and we do not know how those experiences would have helped constitute the (Proustian) SELF.

Recall that, on the conception of the PS, our memories—our MEs—are linked in associative networks, underwritten by vast neuronal circuitries and their connections of varying strength. These networks, in turn, comprise myriad pathways running from node to node, determining the ways in which certain MEs elicit other MEs and those, others, and so forth. But the construction of these pathways is just as unconscious as the integration of the MEs they run between; we may be able to guess, but cannot know precisely why, for the soldier, horse-hoof-prints are more likely to dredge up horseman thoughts than thoughts of the horse-and-carriage ride he and his wife took before he left for the war; or why, for Marcel, his "vast structure of recollection" was folded into the *madeleine* cake instead of some other relic from his childhood. Nor is it necessarily the case that there are always, or even frequently, multiple pathways to the same MEs, and thus less need

for any one ME in particular. So, while we may believe that the erasure of one seemingly inconsequential memory is itself inconsequential to our identities, we can see how such erasure can also have far-reaching effects. If, for example, some ME that would have become a cue is foreclosed from consolidating—say, the taste of Marcel's *madeleine*—then one's SELF has been drastically altered, in the same way that destroying a single of library index can foreclose copious amounts of information. For these reasons, interrupting the process of memory consolidation might pose a greater risk to our SELVES than is initially apparent. Such modifications might not only preclude the ME of the *occurrent* experience from forming, which our brains may have seen fit to integrate into LTM and thereby positively constitute the SELF, but in so doing might also preclude the imminent self-integration of an ME able to cue indefinitely many others, effectively locking indefinitely large stores of memory—big chunks of our selves—in an impenetrable vault.

Of course, the prospect of such memory modification does not risk disintegrating the (Proustian) SELF *tout court*, since a PS dispossessed of certain memories, while perhaps not the *same* PS, is a PS nevertheless. Moreover, for such memory modification to alter the SELF in a manner we might think problematic, it must do more drastically than, or at least sufficiently dissimilarly to, how the PS perpetually modifies itself. In other words, we must be clear on whether our ethical concern regards the PS as such—that is, as an entity which can continue in some form or another, despite having been artificially modified—or whether it regards the individual patient's PS, as a SELF whose uniqueness and diachronicity we should be interested in preserving. While there is undoubtedly more to say on this matter, I think we can at least see how certain of the aforementioned neurobiological theories can provide reasonable answers to these questions. For example, recall that according to reconsolidation theory, the very act of remembering some ME causes that ME to become labile, and therefore able to incorporate concurrently consolidating experiences—

but also, therefore, liable to *disruption* by OMM. A patient undergoing OMM could thus expose herself to losing access, not unlike CR, to a considerable about of MEs simply in virtue of *remembering*—a phenomenon unlikely (if likely at all) to occur in virtue of natural, everyday reconsolidation. In this way, going forward, such a patient risks having lost not *a* PS, but *her* PS; and therefore, insofar as we consider not mere SELF-existence, but SELF-diachronicity, valuable, we might consider OMM potentially harmful.

6 CONCLUSION

Before I offer some concluding remarks, let us take stock of where we've been. We began by examining the not-so-distant prospect of neuromodulation, and why people might be averse to it. Although neuromodulation's potential applications are legion, we homed in on *memory* modification as its paradigm, and optogenetics as its most plausible means. To concretize our analysis, we examined research demonstrating that forgetting meals can cause one to eat more, and more often, than usual, and stipulated therefrom that clinicians might want to use OMM to treat eating disorders like anorexia. We then asked: Might such a therapy realize our (the folk's) misgivings about memory modification? Specifically, might altering our memories somehow alter *who we are*—alter our very selves?

To begin investigating this question, we posited that we must first elucidate the concepts of, and the connections between, "memory" and "self." Thus, my primary aim was to furnish a SELF-conception that both vindicates our intuitive worries about neuromodification and provides a starting point for further ethical analyses. This led us to the conception of the Proustian Self (PS), which posits that not only do our *memories* constitute the SELF, but *how we remember*—the relations between our memories—does as well. Finally, given this plausible conception of the SELF in

relation to memory, we examined the potential effects OMM might visit upon the PS; Namely, because of the neurobiological evidence and theories about the formation, storage, and retrieval of memories, in addition to cases like that of CR, it may be the case that hippocampal manipulations, even if targeting seemingly inconsequential events, can exert far-reaching effects on one's access to information in memory, plausibly resulting in a drastically different SELF than one might have otherwise been.

Further analyses of (optogenetic) memory modification might consider whether, and how, such effects might be morally problematic. Perhaps we might wonder whether the PS is, in fact, a mere psychological construct, and thus whether, irrespective of its phenomenological character, it is morally permissible to alter it in certain ways. Perhaps, that is, we can separate the fact that, via OMM, one's SELF will take a drastically different course that it otherwise would have from the *morality* of modifying memory. Mightn't we want to erase (or otherwise) modify, for example, debilitatingly traumatic memories if we could—the kind that, we think, cannot conceivably be an ingredient in a positive, happy life?

These questions reveal the need for further work on the concept of memory modification and the notion of *harm*. Doing "harm" to the SELF might refer simply to exerting the kind of effect that I have posited OMM exerts on the PS: that which issues in a different SELF—that breaks SELF-diachronicity—irrespective of the net well-being or suffering of the SELF under consideration; or, on the other hand, it might incorporate more complex neuro-ethical aspects such as *authenticity*, *autonomy*, and *informed consent*. In any case, answers to these questions of memory modification's moral landscape, in light of the Proustian Self, will be required to undertake the increasingly complex decision-theory analyses by which we can weigh the risks of OMM and similar therapies

against those of foregoing them—risks not only to lives but to livelihoods; not only to patients but to their *selves*.

REFERENCES

Arnold, C. (2013). *Decoding Anorexia: How breakthroughs in science offer hope for eating disorders*. Routledge.

Besson, L. (2014). *Lucy*. Universal Pictures.

Bingen, B. O., Engels, M. C., Schalij, M. J., Jangsangthong, W., Neshati, Z., Feola, I., Ypey, D. L., Askar, S. F., Panfilov, A. V., & Pijnappels, D. A. (2014). Light-induced termination of spiral wave arrhythmias by optogenetic engineering of atrial cardiomyocytes. *Cardiovascular Research*, *104*(1), 194–205.

Brumberg, J. J. (2000). *Fasting Girls: The history of anorexia nervosa*. Vintage Books.

Burger, N. (2011). *Limitless*. Relativity.

Busskamp, V., Picaud, S., Sahel, J.-A., & Roska, B. (2012). Optogenetic therapy for retinitis pigmentosa. *Gene Therapy*, *19*(2), 169–175.

Cattabeni, F., Gardoni, F., & Di Luca, M. (2004). Molecular biology of postsynaptic structures. In R. W. Davies & B. J. Morris (Eds.), *Molecular biology of the neuron* (2nd ed, pp. 165–180). Oxford University Press.

Cela, E., McFarlan, A. R., Chung, A. J., Wang, T., Chierzi, S., Murai, K. K., & Sjöström, P. J. (2019). An optogenetic kindling model of neocortical epilepsy. *Scientific Reports*, *9*(1), 1–12.

Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.

Cheke, L. G., Simons, J. S., & Clayton, N. S. (2016). Higher body mass index is associated with episodic memory deficits in young adults. *Quarterly Journal of Experimental Psychology*, *69*(11), 2305–2316.

Connolly, C. N. (2004). Protein trafficking in neurons. In R. W. Davies & B. J. Morris (Eds.), *Molecular biology of the neuron* (2nd ed, pp. 75–101). Oxford University Press.

Conway, M. A., & Loveday, C. (2010). Accessing Autobiographical Memories. In J. H. Mace (Ed.), *The Act of Remembering: Toward an Understanding of How We Recall the Past* (pp. 56–70). Wiley-Blackwell.

Cott, J. (2005). *On the Sea of Memory*. Random House.

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? In W. S. Sossin, J.-C. Lacaille, V. F. Castellucci, & S. Belleville (Eds.), *Progress in Brain Research* (Vol. 169, pp. 323–338). Elsevier. https://doi.org/10.1016/S0079-6123(07)00020-9

Crocini, C., Ferrantini, C., Coppini, R., Scardigli, M., Yan, P., Loew, L. M., Smith, G., Cerbai, E., Poggesi, C., & Pavone, F. S. (2016). Optogenetics design of mechanistically-based stimulation patterns for cardiac defibrillation. *Scientific Reports*, *6*, 35628.

Debiec, J., & Nader, K. (2004). Disruption of reconsolidation but not consolidation of auditory fear conditioning by noradrenergic blockade in the amygdala. *Neurscience*, *129*, 267–272.

Deng, C., Yuan, H., & Dai, J. (2018). Behavioral manipulation by optogenetics in the nonhuman primate. *The Neuroscientist*, *24*(5), 526–539.

Dennett, D. C. (1992). The Self as a Center of Narrative Gravity. In *Self and Consciousness: Multiple Perspectives*. Hillsdale, NJ: Erlbaum.

Dierhold, G. H. (1898). The Art of Expression. *Self Culture*, *7*, 235–239.

Diester, I., Kaufman, M. T., Murtaza, M., Pashaie, R., Goo, W., Yizhar, O., Ramakrishnan, C., Deisseroth, K., & Shenoy, K. V. (2011). An optogenetic toolbox designed for primates. *Nature Neuroscience*, *14*(3).

DiGuiseppi, J., & Zuo, J. (2019). The awesome power of optogenetics in hearing research. *Neuroscience Letters*.

Dudai, Y., & Eisenberg, M. (2004). Rites of passage of the engram: Reconsolidation and the lingering consolidation hypothesis. *Neuron*, *44*, 93–100.

Duncan, M. (2019). The self shows up in experience. *Review of Philosophy and Psychology*, *10*(2), 299–318.

Dunkle, E., & Dunkle, C. B. (2015). *Elena Vanishing*. Chronicle Books.

Feliu, N., Neher, E., & Parak, W. J. (2018). Toward an optically controlled brain. *Science*, *359*(6376), 633–634.

Fenno, L., Yizhar, O., & Deisseroth, K. (2011). The development and application of optogenetics. *Annual Review of Neuroscience*, *34*.

Finnie, P. S. B., & Nader, K. (2012). The role of metaplasticity mechanisms in regulating memory stabilization. *Neuroscience and Behavioral Reviews*, *36*, 1667–1707.

Fivush, R., & Waters, T. E. (2019). Development and organization of autobiographical memory form and function. *The Organization and Structure of Autobiographical Memory*, 52.

Forrest, P. (n.d.). The Identity of Indiscernibles. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016).

Funk, C., Kennedy, B., & Sciupac, E. P. (2016). *U.S. Public Wary of Biomedical Technologies to 'Enhance' Human Abilities*. Pew Research Center.

Funk, C., Rainie, L., & Smith, A. (2018). What Worries People about Future Science and Tech Innovations? *Observations*, *2020*.

Gallagher, S. (2011). *The Oxford Handbook of the Self*. Oxford University Press.

Gaub, B. M., Berry, M. H., Holt, A. E., Isacoff, E. Y., & Flannery, J. G. (2015). Optogenetic vision restoration using rhodopsin for enhanced sensitivity. *Molecular Therapy*, *23*(10), 1562–1571.

Gilbert, F., Harris, A. R., & Kapsa, R. (2014). Controlling brain cells with light: Ethical considerations for optogenetic clinical trials. *AJOB Neuroscience*, *5*(3), 3–11.

Giordano, S. (2005). *Understanding Eating Disorders: Conceptual and ethical issues in the treatment of anorexia nervosa and bulimia nervosa*. Oxford University Press.

Goldie, P. (2012). *The Mess Inside: Narrative, Emotion, and the Mind*. Oxford University Press.

Gondry, M. (2004). *Eternal Sunshine of the Spotless Mind*. Focus Features.

Habermas, T., & Bluck, S. (2000). Getting a life: The emergence of the life story in adolescence. *Psychological Bulletin*, *126*(5), 748.

Habermas, T., & Köber, C. (2015). Autobiographical reasoning in life narratives buffers the effect of biographical disruptions on the sense of self-continuity. *Memory*, *23*(5), 664–674.

Habermas, T., Negele, A., & Mayer, F. B. (2010). "Honey, you're jumping about"—Mothers' scaffolding of their children's and adolescents' life narration. *Cognitive Development*, *25*(4), 339–351.

Hannapel, R., Henderson, Y. H., Nalloor, R., Vazdarjanova, A., & Parent, M. B. (2017). Ventral hippocampal neurons inhibit postprandial energy intake. *Hippocampus*, *27*(3), 274–284.

Hannapel, R., Ramesh, J., Ross, A., LaLumiere, R. T., Roseberry, A. G., & Parent, M. B. (2019). Postmeal optogenetic inhibition of dorsal or ventral hippocampal pyramidal neurons increases future intake. *ENeuro*.

Hebben, N., Corkin, S., Eichenbaum, H., & Shedlack, K. (1985). Diminished ability to interpret and report internal states after bilateral medial temporal resection: Case HM. *Behavioral Neuroscience*, *99*(6), 1031.

Henderson, Y. H., Gerard, P., & Parent, M. B. (2013). Hippocampal neurons inhibit meal onset. *Hippocampus*, *23*(1), 100–107.

Higgs, S. (2008). Cognitive influences on food intake: The effects of manipulating memory for recent eating. *Physiology & Behavior*, *94*(5), 734–739.

Higgs, S. (2016). Cognitive processing of food rewards. *Appetite*, *104*, 10–17.

Holmes, R. (2008). A meander through memory and forgetting. In A. S. Byatt & H. H. Wood (Eds.), *Memory: An anthology* (p. 99). Chatto & Windus.

Huang, F., Tang, B., & Jiang, H. (2012). Optogenetic investigation of neuropsychiatric diseases. *International Journal of Neuroscience*, *123*(1), 7–16.

Hume, D. (1888). *Treatise of Human Nature* (L. A. Selby-Bigge, Ed.). Clarendon Press.

Hutto, D. (2008). *Folk Psychological Narratives*. Bradford Books.

Jones, A., & Brooker, Charlie. (2011). Black Mirror [Television series]. In *Black Mirror*. Endemol Shine UK.

Kravitz, A. V., Freeze, B. S., Parker, P. R., Kay, K., Thwin, M. T., Deisseroth, K., & Kreitzer, A. C. (2010). Regulation of parkinsonian motor behaviours by optogenetic control of basal ganglia circuitry. *Nature*, *466*(7306), 622–626.

Landy, J. (2001). "Les Moi en Moi": The Proustian Self in Philosophical Perspective. *New Literary History*, *32*(1), 91–132.

Lee, J. L., DiCiano, P., Thomas, K. L., & Everitt, B. J. (2005). Disrupting reconsolidation of drug memories reduces cocaine-seeking behavior. *Neuron*, *47*, 795–801.

Lee, J. L., Milton, A. L., & Everitt, B. J. (2006). Cue-induced cocaine seeking and relapse are reduced by disruption of drug memory reconsolidation. *Journal of Neuroscience*, *26*, 5881–5887.

Levitan, I. B., & Kaczmarek, L. K. (2015). *The Neuron: Cell and Molecular Biology* (4th ed.). Oxford University Press.

Locke, J. (1979). *An Essay Concerning Human Understanding* (P. H. Nidditch, Ed.). Oxford University Press.

MacIntyre, A. (1984). *After Virtue* (2nd ed.). University of Notre Dame Press.

Malcolm, N. (1971). *Problems of Mind: Descartes to Wittgenstein*. Harper Torchbooks.

Mandelbaum, E. (2017). *Associationist Theories of Thought*. Stanford Encyclopedia of Philosophy. .

Manning, M. (1996). *Undercurrents*. HarperOne.

Martin, A., Davidson, T., & McCrory, M. (2018). Deficits in episodic memory are related to uncontrolled eating in a sample of healthy adults. *Appetite*, *124*, 33–42.

McKenzie, S., & Eichenbaum, H. (2011). Consolidation and reconsolidation: Two lives of memories? *Neuron*, *71*, 224–233.

McLaughlin, B., & Bennett, K. (2018). Supervenience. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2018). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2018/entries/supervenience/

Moser, T. (2015). Optogenetic stimulation of the auditory pathway for research and future prosthetics. *Current Opinion in Neurobiology*, *34*, 29–36.

Nader, K. (2003). Memory traces unbound. *Trends in Neurosciences*, *26*, 65–72.

Nader, K., Schafe, G. E., & LeDoux, J. E. (2000). Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature*, *406*, 722–726.

Nahmias, E. (2011). Intuitions about free will, determinism, and bypassing. In R. Kane (Ed.), *The Oxford Handbook of Free Will* (2nd ed., pp. 555–576). Oxford University Press.

Nehring, C. (2005). *Heloise & Abelard: Love Hurts*. The New York Times. https://www.nytimes.com/2005/02/13/books/review/heloise-abelard-love-hurts.html

Nelson, K. (2003). Narrative and the Emergence of a Consciousness of Self. In G. D. Fireman, T. E. McVay Jr., & O. J. Flanagan (Eds.), *Narrative and Consciousness*. Oxford University Press.

Nolan, C. (2010). *Inception*. Warner Bros. Pictures.

Pama, E., Colzato, L. S., & Hommel, B. (2013). Optogenetics as a neuromodulation tool in cognitive neuroscience. *Frontiers in Psychology*, *4*, 610.

Papanicolaou, A. C. (2006). *The Amnesias: A Clinical Textbook of Memory Disorders*. Oxford University Press.

Paralikar, K., Cong, P., Yizhar, O., Fenno, L. E., Santa, W., Nielsen, C., Dinsmoor, D., Hocken, B., Munns, G. O., & Giftakis, J. (2010). An implantable optical stimulation delivery system for actuating an excitable biosubstrate. *IEEE Journal of Solid-State Circuits*, *46*(1), 321–332.

Parent, M. B. (2016a). Cognitive control of meal onset and meal size: Role of dorsal hippocampal-dependent episodic memory. *Physiology & Behavior*, *162*, 112–119.

Parent, M. B. (2016b). Dorsal hippocampal–dependent episodic memory inhibits eating. *Current Directions in Psychological Science*, *25*(6), 461–466.

Parent, M. B., Darling, J. N., & Henderson, Y. H. (2014). Remembering to eat: Hippocampal regulation of meal onset. *American Journal of Physiology: Regulatory Integrative and Comparative Physiology*, *306*(10), R701–R713.

Parfit, D. (1984). *Reasons and Persons*. Oxford University Press.

Pedersen, N. P., & Gross, R. E. (2018). Neuromodulation Using Optogenetics and Related Technologies. In *Neuromodulation* (pp. 487–500). Elsevier.

Pfister, W. (2014). *Transcendence*. Warner Bros. Pictures.

Pitman, R. K. (2011). Will reconsolidation blockade offer a novel treatment for post-traumatic stress disorder? *Fronteirs of Behavioral Neuroscience*, *5*, 11.

Pope, A. (1717). *Eloisa to Abelard*. The Poetry Foundation. https://www.poetryfoundation.org/poems/44892/eloisa-to-abelard

President's Council on Bioethics. (2010). Memory Blunting: Ethical Analysis. In M. J. Farah (Ed.), *Neuroethics: An introduction with readings* (pp. 88–97). The MIT Press.

Proust, M. (1913). *A la recherche du temps perdu: Vol. I* (C. K. S. Moncrieff, Trans.). Random House, Inc.

Przybyslawski, J., Roullet, P., & Sara, S. J. (1999). Attenuation of emotional and non-emotional memories after their reactivation: Role of beta adrenergic receptors. *Journal of Neuroscience*, *19*, 6623–6238.

Ricoeur, P. (1994). *Oneself as Another* (K. Blamey, Trans.). Chicago University Press.

Robinson, E., Aveyard, P., Daley, A., Jolly, K., Lewis, A., Lycett, D., & Higgs, S. (2013). Eating attentively: A systematic review and meta-analysis of the effect of food intake memory and awareness on eating. *The American Journal of Clinical Nutrition*, *97*(4), 728–742.

Rowlands, M. (2017). *Memory and the Self: Phenomenology, Science, and Autobiography*. Oxford University Press.

Rozin, P., Dow, S., Moscovitch, M., & Rajaram, S. (1998). What causes humans to begin and end a meal? A role for memory for what has been eaten, as evidenced by a study of multiple meal eating in amnesic patients. *Psychological Science*, *9*(5), 392–396.

Rudy, J. W. (2018). *The Neurobiology of Learning and Memory* (2nd ed.). Oxford University Press.

Sacks, O. (2007, September 24). The Abyss: Music and amnesia. *The New Yorker*. https://www.newyorker.com/magazine/2007/09/24/the-abyss

Schechtman, M. (1996). *The Constitution of Selves*. Cornell University Press.

Schechtman, M. (2011). The Narrative Self. In S. Gallagher (Ed.), *The Oxford Handbook of the Self* (pp. 394–416). Oxford University Press.

Siderits, M. (2011). Buddhist Non-Self. In S. Gallagher (Ed.), *The Oxford Handbook of the Self* (pp. 298–315). Oxford University Press.

Spinoza, B. (2008). Ethics. In H. H. Wood & A. S. Byatt (Eds.), *Memory: An Anthology* (p. 169). Chatto & Windus.

Strawson, G. (2009). *Selves: An Essay in Revisionary Metaphysics*. Oxford University Press.

Strawson, G. (2011). The Minimal Subject. In S. Gallagher (Ed.), *The Oxford Handbook of the Self* (pp. 253–278). Oxford University Press.

Suárez, L. D., Smal, L., & Delorenzi, A. (2010). Updating contextual information during consolidation as result of a new memory trace. *Neurobiology of Learning and Memory*, *93*(4), 561–571.

Sweeny, C. (2015). Limitless [Television series]. In *Limitless*. CBS Television Distribution.

Taylor, C. (1989). *Sources of the Self*. Harvard University Press.

Thackeray, W. M. (1889). *The Newcomes: Memoirs of a Most Respectable Family* (Vol. 1). Houghton, Mifflin & Co.

Tonegawa, S., Morrissey, M. D., & Kitamura, T. (2018). The role of engram cells in the systems consolidation of memory. *Nature Reviews Neuroscience*, *19*(8), 485–498.

Touriño, C., Eban-Rothschild, A., & de Lecea, L. (2013). Optogenetics in psychiatric diseases. *Current Opinion in Neurobiology*, *23*(3), 430–435.

Tronson, N. C., Wiseman, S. L., Olausson, P., & Taylor, J. R. (2006). Bidirectional behavioral plasticity of memory reconsolidation depends on amygdalar protein kinase A. *Nature Neuroscience*, *2*, 161–169.

Van Dyke, H. (1895). *Little Rivers: A Book of Essays in Profitable Idleness*. Charles Scribner's Sons.

Velleman, J. D. (2006). The Self as Narrator. In *Self to Self: Selected Essays* (pp. 203–223). Cambridge University Press.

Verhoeven, P. (1990). *Total Recall*. TriStar Pictures.

West, A. (2014). *Manipulating Memory through Optogenetics: A Conversation with Neuroscientists Xu Liu and Steve Ramirez*. Behavioral Scientist. https://behavioralscientist.org/manipulating-memory-through-optogenetics-qa-with-neuroscientists-xu-liu-and-steve-ramirez/

Wiseman, L. (2012). *Total Recall*. Sony Pictures Releasing.

Yazdan-Shahmorad, A., Silversmith, D. B., Kharazia, V., & Sabes, P. N. (2018). Targeted cortical reorganization using optogenetics in non-human primates. *Elife*, *7*, e31034.

Zahavi, D. (2011). Unity of Consciousness and the Problem of Self. In S. Gallagher (Ed.), *The Oxford Handbook of the Self* (pp. 316–335). Oxford University Press.

Zhang, T. R., Larosa, A., Di Raddo, M.-E., Wong, V., Wong, A. S., & Wong, T. P. (2019). Negative memory engrams in the hippocampus enhance the susceptibility to chronic social defeat stress. *Journal of Neuroscience*, *39*(38), 7576–7590.