Knowing (From) Me, Knowing (From) You: Essays on Memory and Testimony

Chloe Wall

A thesis submitted for the degree of
Doctor of Philosophy
at the University of Otago
Ōtepoti Dunedin, Aotearoa New Zealand

"No one knows anything, really. It's all rented, or borrowed."

-Ian McEwan, *Saturday*

ABSTRACT

This thesis is a collection of four related but self-standing chapters about memory and testimony. In Chapter 1, I argue that memory and testimony are analogous because both are reconstructive, incorporating information from sources in relevantly similar ways. In Chapter 2, I begin with the standard taxonomy of memory, according to which memory-how (procedural memory) is distinct from memory-that (episodic or semantic memory). From there, I develop an account of testimony-how by arguing that testimony need not be propositional. In Chapter 3, I turn to the curious case of the Mandela Effect and argue that it is an instance of collective confabulation, in which large groups of people develop highly similar apparent memories of events that never occurred. For at least some of these cases, I claim, testimony is an integral ingredient in the production of collective confabulations. In Chapter 4, I proceed from the analogy between testimony and memory and argue that testimonial injustice has a memorial analogue, which I call *memorial injustice*. I consider internalised false confessions to be an example of memorial injustice, and I identify failures of metacognition as being a key component the precipitation thereof. Ultimately, the work in this thesis leads me to the conclusion that while social epistemologists have focussed on the social factors influencing the epistemology of testimony, the ways that social phenomena influence memory and remembering are too great to ignore. I do not offer a positive account of how social epistemology ought to treat memory, but I hope that thinking about the relationship and similarities between memory and testimony offers a new perspective from which to view both.

ACKNOWLEDGMENTS

No research takes place in a vacuum, unless you work at CERN. I don't work at CERN, though, and so producing this thesis has involved innumerable contributions from those around me, and I would be remiss not to acknowledge the philosophical, professional, and personal support that I have received over the past years.

I am, first and foremost, grateful to my supervisors. I owe a tremendous intellectual and professional debt to Kirk Michaelian for his willingness to take me on as a student, his faith in my project, his encyclopaedic knowledge of the philosophy of memory, and his generosity of time and feedback. His intellectual fingerprints on this thesis will be evident to anyone familiar with his work. Thank you also to Fabien Medvecky, who joined my committee in the early stages of the project and brought with him a great sense of humour and an infectious enthusiasm for epistemic in/justice that fuelled me especially toward the end. And finally, thank you to Mike LeBuffe, an exceptionally perspicacious philosopher with something useful to say about nearly everything, for inheriting me and my project with grace, unfailing kindness, and steadfast support. All three have been endless sources of guidance, encouragement, and advice, and I count myself exceptionally lucky to have had them as supervisors and mentors while producing this thesis.

The academic support did not begin and end with my supervisors. Throughout the production of this thesis, I have met and enjoyed the intellectual and social company of philosophers from further afield than Ōtepoti Dunedin. In reverse alphabetical order, because (i) I can, and (ii) those of us with names toward the end of the alphabet have to stick together, those philosophers are Joe Ulatowski, John Sutton, Sarah Robins, Chris McCarroll, Justine Kingsbury, Jordi Fernández, Carl Craver, and Hannah Clark-Younger. Many of them proved fantastic conversation (or verbal sparring) partners at the pub, between conference presentations, biking up Big Easy, or over email, and their ideas have inevitably bled into mine.

I have benefited from the friendship and philosophical contributions from many of the postgraduate students whom I have been fortunate to call my colleagues and friends. A few deserve explicit mentions (also in reverse alphabetical order). Zach "Sprinklehurst" Swindlehurst always had time for a bad coffee or "business meeting" and to complain about something or other having to do with normativity, and also generously agreed to proofread much of this thesis. (Any remaining errors are my own.) André Sant'Anna read and always provided incisive and constructive comments on several drafts of the material in this thesis. Manuel Lechthaler, the best officemate anyone could ever ask for, was always keen to grab a coffee, to speculate about memory and mereology, and to argue with me about whether Superman and Clark Kent are identical. (My view, which I present without argument, is that they are not.) It is appropriate that Jon "JP" Keyzer is a utilitarian (WRONG!), because he has certainly maximized the good in my life, with Brekkie Club, bikes, beers, video games, La Croix, $3 lunch, Bad Movie Bingo, and lots and lots of arguing. And Briony Blackmore is an exceptionally generous and thoughtful person with a quiet insistence on discovering what is really going on underneath the surface. She is, to me, a model of discipline and perseverance, and is the best watermelon-sharing companion around. Each is one of a kind, and to say that I have benefited from their friendship and philosophical contributions, as I did above, is to fail to capture and to honour the close friendships we have cultivated. They and all the members of the Department of Philosophy at the University of Otago have contributed in immeasurable ways to the development of the ideas herein. I also wish to thank my non-philosophy friends in Aotearoa New Zealand, including Avery Underwood, Kate Stevens, Matthew Schep, Adam Rance, Emer Lyons, Dan Jaffe, Anita Gibbs, Charlie Campbell, Hahna Briggs, and the staff of Eureka.

I was lucky to have support from both sides of the Pacific Ocean. From Canada, my mother Sarah, my father Dan, my brother Adam, my sister Olivia, and her husband Blake have always been there for me. Their support has taken myriad

forms, including pep talks, working through my ideas, sending photos of the dogs Watson, Tori, and Baxter, and quoting lines from our favourite movies. I have thought of you every day of my time here, and really could not have done this without you. Other dear Canadian friends who often swooped in on rainy days when I needed it the most are Karen Simons, Patrick Salmers, Celyne Runzer, Allison Preyde, and Erin McCarty-Buijs.

Three families in Aotearoa New Zealand welcomed me with open arms. To Melanie Beres, Louise Pearman-Beres, Ovita Beres, Boo, and the rest of the menagerie, thank you for all you have done for me. There is simply no way I can repay you for how welcoming, loving, and supportive you have been to me during my time in Ōtepoti Dunedin, for all things personal and academic. Everybody needs an auntie to steer them through life, and I am lucky enough to have found two. To Alexa Anderson, Hamish Cartwright, and Pippin, thank you allowing me to join your household (more than once!), and for your friendship and generosity from, quite literally, my first day in town. To Simon Lovatt and Amy Van Wey Lovatt, thank you for welcoming me—an actual stranger—into your home, and for your generosity, hospitality, willingness to chat about philosophy, and cheerleading (and much-needed nagging!) at the very end. I am so very grateful for your continued friendship.

Completion of this thesis was made possible by the generous funding I received from the Commonwealth Scholarship, administered by the University of Otago. I also received funding from the Australasian Association of Philosophy, the New Zealand Association of Philosophers, the Otago Institute for the Arts and Sciences, the Humanities Division at the University of Otago, and Sven Bernecker's Alexander von Humboldt Professorship award to present the work in this thesis in Aotearoa New Zealand and abroad; I am grateful for the generosity of these bodies. I am also grateful to six anonymous journal referees and to the audiences at the many conferences, symposia, and workshops I was privileged enough to attend for their

feedback on the work in this thesis throughout its development. In particular, I would like to acknowledge Jordi Fernández for comments that greatly influenced the development of Chapter 1, to Hannah Clark-Younger for helpful comments on Chapter 2, to James McKeahnie for proofreading Chapter 2, to Celia Harris for feedback on a conference presentation corresponding to Chapter 3, and to Carl Craver for interesting discussions about epistemic injustice and memory. Portions of this thesis were also submitted for journal publication in different forms, and have benefited from the comments I received from anonymous reviewers.

It is a testament to the intellects and characters of the people around me that these acknowledgments have been so difficult to write: there are simply so many people who have contributed to the production of this thesis that I would have had to start writing them down on Day 1 of the degree to keep track of them all. Being a memory researcher does not, unfortunately, imbue me with a good memory, and I will undoubtedly have neglected to mention some people who have provided support, feedback, a listening ear, or gin, and in doing so made completing this project just a little easier (or at least a bit more fun). *Mea culpa*, and may you all be realised in many possible worlds.

CONTENTS

INTRODUCTION

Some of my earliest memories come from a trip my family took to my father's hometown of St. John's, a comparatively small city on the east coast of Canada. Having grown up in a landlocked province, my sister and I had never seen the ocean before, and our parents took us on a whale-watching trip on a boat called the *Scadamia*.

Children at young ages seem to delight in being upside-down, and I was no exception. Aboard the *Scadamia*, I sat on a bench and bent over until I was upside-down, looking between my legs through the railing out at the water. While dangling off the bench, I noticed an earring on the deck of the boat, and recognised it as my mother's. I sat upright, looked at her ears, and saw that an earring was missing. I retrieved the fallen earring and gave it back to her, feeling proud of myself for having saved the day and glad that she was happy to have the earring back. (She hadn't, in fact, noticed that it was gone.)

The memory is completely innocuous in almost every regard, but it stands out in my mind for this reason: when recounting it to a family friend many years later, my sister interrupted me and said, "*You* didn't find Mom's earring – *I* did! I remember hanging upside-down and seeing the earring on the ground." My (our?) memory of rescuing the earring aboard the *Scadamia* is what is known in the memory literature as a disputed memory, in which two people (often, but not always, twins) claim ownership of the same memory (A. S. Brown et al. 2015; Ikier et al. 2003; Küntay, Gülgöz, and Tekcan 2004). Both my sister and I genuinely believe ourselves to be the protagonist of the story. When we consulted our mother to settle the dispute, she was unable to remember what had happened, and so the truth has been lost to the past.

It is likely that my sister, my mother, and I will never know who really rescued Mom's earring aboard the *Scadamia*. I can, of course, insist that the memory

is mine, but then again, my sister can do exactly the same, and with no testimony from my mother to break the tie, we are at an epistemic impasse. Our knowledge of who rescued the earring is limited by what we remember and how we tell each other about it. In the end, it probably does not matter who really rescued the earring, but plenty of epistemic impasses *do* have great consequences for those caught within. Thinking about the relationship between memory and testimony, and the epistemic positions they allow us to take, is both the purpose and theme of this thesis.

I begin this project with a claim that I hope will not be too controversial: that memory and testimony alike are indispensable to our successful functioning as agents. Philosophers have frequently remarked on the similarities between memory and testimony. Dummett (1994), for instance, has remarked that memory may be thought of as the testimony of the past self. Sosa (1991, 218) has said that "memory is a psychological mechanism that conveys beliefs across stages of [one individual's] life," and testimony is "a social mechanism that conveys beliefs across lives at a time." Testimony and memory are also similar in that they are both vital to our successful epistemic functioning. If we were to remove all those beliefs formed purely on the word of others, or all those beliefs for which our only evidence lies in memory, we would be left with precious little epistemic ground on which to stand. The aim of this project is to explore the similarities, relationships, and interactions between memory and testimony with respect to their epistemic function. This thesis is a collection of four somewhat self-standing chapters that approach this question from different perspectives.

Chapter 1 proceeds from the observation that the tendency in the epistemologies of memory and testimony has been to assume that the two are analogous. Crucially, support for this analogy comes from the supposition that both memory and testimony play primarily transmissive roles: memory, the claim goes, transmits content within a subject from an earlier time to a later time, while testimony transmits content from one subject to another. Given the lack of empirical

support for the conceptions of memory and testimony that ground the transmissive analogy, however, the pervasiveness of the analogy is curious at best and indefensible at worst. But the problem is not only that there is no empirical support for claims that memory and testimony are transmissive; the problem is that empirical evidence points in another direction altogether by showing that memory and testimony are matters of construction and reconstruction. I have two aims in this chapter. First, I attack the transmissive analogy and show that it is threatened by evidence about memory and communication. Second, I show that the analogy turns out to be correct, as long as we accommodate—even capitalise on—the (re)constructive features of memory and testimony. I sketch the general features of this constructive analogy and finish by considering a few worries about its limits. I return to the features of the analogy in Chapters 3 and 4.

Before proceeding to the implications of the analogy for our thinking about memory and testimony, the next two chapters tackle issues that put tension on the analogy, but lie specifically with testimony (Chapter 2) and memory (Chapter 3), respectively. Thinking about the analogy between memory and testimony invites many questions about how far the analogy goes, In Chapter 2, I begin with the standard taxonomy of memory (Squire 2004), which divides memory into two kinds: declarative memory, including memory for facts and events, and non-declarative memory, including procedural memory, or memory for how to do things. One point at which the analogy between memory and testimony might appear to break is at the highest level of the memory taxonomy. Procedural memory, in particular, appears to pose a problem for the analogy, since it is unclear how testimony might convey *how* one does something.

This is especially problematic given that received wisdom has it that the currency of testimony is propositions. Though the literature is saturated with disagreement over the manifestation of the proposition (e.g., belief vs. statement vs. knowledge vs. something else altogether) and the process it undergoes during

testifying (e.g., replication vs. transmission vs. generation), the propositionality of testimony is virtually universally accepted. This is often taken to preclude certain types of utterances—namely, interrogatives and imperatives—from qualifying as testimony, since they seem to lack truth conditionality, a defining feature of propositions. As a result, normally only sentences in the declarative mood qualify as testimony. In Chapter 2, I focus on imperatives. Specifically, I am interested in instructions, which are epistemologically interesting if understood as attempts to instil beliefs in an instructee. Under normal declarative circumstances, it would be entirely natural to think of instilling a belief in another (or attempting to do so) as an instance of testimony; what complicates matters in the case of instructions is that they are seemingly non-propositional. The aforementioned propositionality condition in testimony precludes instructions from qualifying as testimony, and yet, instructions still seem to fill much the same functional niche that paradigmatic testimony does. If, in any other case, we would not hesitate to apply the label "testimony" to the transfer of knowledge from one person to another, it is puzzling that we should be disinclined to do so in the case of instructions.

Chapter 2, then, proceeds from a puzzle in reconciling three inconsistent claims arising from the above considerations: first, that some instructions are delivered in the imperative mood (i.e. in a non-propositional way), second, that instructions are a means to testify as to how to do something, and third, that testimony is propositional. I ultimately argue that the way to resolve the inconsistent triad is to reject the claim that testimony is always propositional, and conclude that while typical testimonial utterances can be characterised as testimony-that, instructions amount to testimony-how. As a result, we ought to develop epistemologies of testimony that can account for propositional and non-propositional forms of testimony alike. Finally, I argue that doing so strengthens the analogy between memory and testimony that I defended in Chapter 1, as it shows that the analogy applies not only to memory-that and propositional testimony

(testimony-that), but that some version of the analogy will apply to non-declarative memory (or memory-how) and testimony-how.

In Chapter 3, which draws on work co-authored with Kourken Michaelian, I enter the woolly world of memory in a communicative (and testimony-filled) space: the Internet. In recent years, popular fora, particularly Reddit, have seen lively discussion of the "Mandela Effect". The Mandela Effect—so called in reference to the paradigm case of a widely shared memory of Nelson Mandela dying in prison in the 1980s—occurs when individuals who have never met each other in person develop highly similar memories of events that never occurred. Popular explanations of this phenomenon (e.g., that the apparently false memories in question are in fact accurate memories of events that occurred in parallel universes) are fanciful, and the scientific literature so far contains no discussion of the effect or the mechanisms giving rise to it. The purposes of this chapter are, first, to make a case for the existence of the Mandela Effect as a novel memory error worthy of scientific attention and, second, to sketch a general account of a mechanism that might give rise to it. My hypothesis is that the Mandela Effect is an instance of collective confabulation. I argue that, given either the causal account of mnemic confabulation defended by Robins (2016a; 2017) and Bernecker (2017) or the reliability account defended by Michaelian (2016b), the effect amounts to confabulation on the collective level. It does not, however, reduce to individual confabulation. Instead, the effect occurs when ordinary misremembering goes online. The interaction among misrememberers results not in correction by peers, as it typically would offline, but in reinforcement of the mismemory by confirmatory testimony and corroboration by others. The Mandela Effect thus amounts to a novel, genuinely collective form of confabulation.

In Chapter 4, I return to the core concern of the interaction between memory and testimony. I introduce Miranda Fricker's (2007) account of epistemic injustice and I argue that, based on my claims in Chapter 1, if memory and testimony are analogous, and there is a form of epistemic injustice specific to testimony, then there

is an analogous form of epistemic injustice for memory. I take internalised false confessions—in which an innocent suspect confesses to a crime he did not commit because hostile, guilt-presumptive, and manipulative interrogation results in the suspect coming to believe that he is guilty—to exemplify this phenomenon. From there, I develop an account of *memorial injustice* by appealing to the role of source monitoring in memory and likening it to mindreading in testimony. I conclude that memorial injustice comes about owing to failures of source monitoring and self-assigned credibility deficits.

Chapters 3 and 4 especially emphasise the interaction of testimony with memory: they show that testimony from outside can influence, modify, and even produce memories within a subject. If this is right, then this has significant implications for epistemology. Epistemologists have traditionally understood memory to be uniquely internal and individual in a way that even perception and reason are not, and this assumption has infused much of the inquiry into the nature of memory and memory justification. What I claim in this thesis directly challenges this assumption. Testimony leaves its fingerprints all over the memorial looking-glass, and to disregard this fact is to deny the myriad ways in which our memories are not entirely internal or individual after all. Let us now venture into the worlds of memory, and testimony, and the smudged looking-glass between them.

CHAPTER 1:
ARE MEMORY AND TESTIMONY ANALOGOUS?

1.1. Introduction

In epistemology, there is a longstanding assumption, made explicit in Dummett's (1994, 252) claim that "memory may be said to be the testimony of one's past self," that memory and testimony are analogous. Though much work in epistemology of memory and testimony has depended on the apparent similarity between the two—namely, that memory and testimony appear to preserve or transmit content (Lackey 2006c)—recent memory research has repeatedly shown that memory is a capacity that *generates* rather than preserves content (see Michaelian (2011c) for a systematic review). If the analogy between memory and testimony hinges on the false claim that memory preserves content, then defenders of the analogy have been barking up the wrong tree all along. Even though the analogy has been fruitful in developing increasingly sophisticated accounts of memory and testimony, we simply cannot endorse it if it is false. At best, what we have come to believe about memory and testimony on the basis of this analogy is true, but only luckily so. At worst, what we have come to believe is false, and needs radical reconsideration.

It is easy to jump to the conclusion that the analogy is beyond redemption, since it rests upon qualities that have turned out not to be so similar after all. But while it is true that the transmissive analogy is mistaken, it turns out that we are not wrong to claim that there is an analogy. Testimony, as we shall see, is similarly constructive in the way that memory is, and so the claim that there is an analogy is correct after all; it will, however, require some rethinking. My arguments in this chapter are first, that testimony is constructive, and second, that memory and testimony really are analogous in virtue of their constructive features. Crucially, I claim, in both remembering and testifying, it is not merely the past self or the testifier who supplies information that is integrated into the final message. Neither does the present self or the recipient get off scot-free: both must play their respective

roles in reconstruction and interpretation. Rather, memory and testimony are both constructive processes that integrate information from a variety of sources to produce coherent messages, which the receiver must decode.

1.2. The transmissive analogy

Testimony and memory are both critical for our successful functioning as social and epistemic agents, but they also exhibit epistemological qualities that motivate thinkers such as Dummett (1994, 264) to defend the "almost exact analogy between memory and testimony." On his view, the analogy between memory and testimony is such that since we are justified in accepting beliefs from memory, then we are similarly justified with respect to testimony.

Two linked similarities between memory and testimony point to this analogy. First, memory and testimony preserve or transmit knowledge: memory is "the retention of knowledge previously acquired" (Dummett 1994, 264), while testimony is "the transmission from one individual to another of knowledge acquired by whatever means" (Dummett 1994, 264). This claim aligns with the *Archival View of Memory* (AVM), the pervasive and enduring historical view of memory as a faculty that creates, stores, and retrieves discrete representations of past events (Robins 2016a). Some version of the view goes back at least to Plato, who described memory as a wax tablet upon which impressions of perceptions are imprinted (*Theaetetus* 191c-d), although it has enjoyed several incarnations in metaphors of memory as palace, library, computer hard drive, and so on (see Brockmeier (2015, chap. 3), Koriat and Goldsmith (1996), and Roediger (1980) for methodical reviews of these metaphors). Traditionally, perception and reason have been taken to generate new knowledge in virtue, *inter alia*, of their capacity to generate new beliefs (e.g., Graham 2006). But unlike perception and reason, neither memory nor testimony have been taken to generate new beliefs. Rather, both *convey* content from a source to a

receiver[1]: memory maintains beliefs over time, while testimony allows one person to gain a belief from another, or so the assumption goes (Lackey 2008).

The second similarity between memory and testimony that motivates Dummett's analogy is the claim that a subject must know that *p* in order to preserve or transmit this knowledge; if it is not known, the subject cannot preserve or transmit knowledge. In Audi's (2006, 43) words, "testimony can increase the number of *knowers* in the world, [but] it cannot increase the number of *propositions known*" (emphasis in original); similarly, Dummett claims, a subject cannot remember something she does not know. This latter claim goes some way toward explaining why neither memory nor testimony generate knowledge.

The analogy defended by Dummett relies not only on implicit assumptions about memory, but about testimony, which he calls "the transmission from one individual to another of knowledge acquired by whatever means" (Dummett 1994, 264); such a characterisation requires that every participant in a testimonial chain know that *p* in order to transmit it. But this view has not gone uncriticised: a later participant in a testimonial chain could come to know that *p* from the first participant even if the intermediate participants fail to know that *p* (Burge 1993; Gelfert 2014; Lackey 1999). On either Dummett's or Lackey's views, however, the first person in the chain must know that *p* through non-testimonial means, such as reason or perception. Although there is disagreement over precisely *what* is being transmitted by testimony—whether it is knowledge, justification, or belief—accounts converge on the point that some act of communication functions as a vehicle for propositions to travel from testifier to recipient. Lehrer calls this view, which is

[1] A brief terminological note is necessary here. Throughout this chapter—and the rest of this thesis—I use the terms "source" and "receiver" to refer to the participants in a generic case of content transmission (whether testimony or memory), where the source asserts that *p* and the receiver gains the belief that *p*. Specifically for memory, I use the terms "past self" or "earlier self" to refer to the source, in contrast with "present self" or "later self" to refer to the receiver. For testimony, I use the terms "testifier" to refer to the source and "recipient" to refer to the receiver.

widely accepted among philosophers (Graham 2006; Lackey 2006a) "the causal theory of knowledge" or "the transmission theory of knowledge" (Lehrer 2006, 145).

This transmission theory aligns with the Code Model of Communication (CMC), an application of Shannon and Weaver's (1949) General Communications Model, developed to be applicable to communication, broadly construed (Turnbull 2003). As Turnbull (2003) characterises CMC, a source encodes information from a message into a signal, and transmits it across a channel, which also contains "noise" (competing signals that originate in some other source). The signal—which may be different from the one sent, depending on noise in the channel and how it has overwritten the original signal—is received at the destination and decoded to receive the message. The picture CMC paints is clear: while it allows that noise-type signals may influence how the signal is received at the destination, it implies that the message decoded by the receiver is the same as the one encoded by the source. At most, it permits information loss, but not the integration of new information. It is not difficult to identify the parallels between this picture and Dummett's characterisation of testimonial chains. The key features are these: at least one participant has some message to pass to the next participant, that message remains largely unchanged throughout the transfer, and the recipient does some work—namely, decoding or interpreting—upon receiving the message.

Standard accounts of testimony have been heavily influenced by the apparent similarities between memory and testimony (Lackey 2006b), so CMC's resemblance to AVM is both obvious and unsurprising. On an AVM-compatible account of memory, an experience is converted into a representation that is preserved until a later point, at which it is retrieved and remains relatively faithful to the original experience; on a CMC-compatible account of testimony, a testifier converts a message into intelligible signals and sends it to a recipient, who receives the signals and gains a message that is, again, largely unchanged from the original message the testifier intended to communicate. In neither case, it is claimed, is new content

generated: in testimony, Goldberg (2001, 517) argues that "the proposition believed on the basis of accepting [a] piece of testimony (the 'received proposition') must derive from the proposition attested to in such a way that the received proposition contains no more information than the one attested to," while in memory, Bernecker (2008; 2010) argues that what is remembered cannot exceed the content of the original experience. It is, however, acknowledged that content might be lost, as in the case of ordinary forgetting or understanding "It is raining" from the utterance "It is raining very hard."[2]; in cases such as these, we find content loss. Furthermore, both AVM and CMC posit a causal connection between the original experience or message (what is encoded and sent by the source) and the representation formed downstream (what is received by the receiver, whether that be the testimonial recipient or the later self). It is, therefore, not difficult to understand why philosophers like Dummett readily endorse a memory–testimony analogy.

1.3. Why the analogy fails: memory is constructive

The family resemblance between AVM and CMC is strong enough that a memory–testimony analogy is intuitively appealing. Both memory and testimony appear to function as a kind of conveyor belt, moving content from one person to another or across time. Illustrative though they may be, the metaphors we use to describe memory are inadequate descriptions of the processes unfolding during remembering. It is curious that despite strong counterevidence from well-documented memory errors, such as misremembering and confabulating,[3] the claim that memory is strictly preservative has endured. This is especially surprising given that not even the proponents of causal theories claim that memory preserves perfectly. Dummett (1994, 265), for instance, admits that "knowledge, like

[2] This example is borrowed from Goldberg (2001).

[3] See De Brigard (2014) and Robins (2016a) for overviews of this literature.

everything else, is liable to be corrupted in transmission,"[4] while Martin and Deutscher (1966, 191) advance the (underdeveloped) idea that "it might seem possible that the memory trace should itself be a suggestible state." As Robins (2016a) points out, however, the problem is not merely that AVM cannot explain memory errors, but rather that the frequency of these errors far exceeds what we would normally expect if AVM were more or less correct. Hence, we cannot justifiably retain confidence in an analogy that hinges on the preservative character of memory, when there is mounting evidence that this view is mistaken. Memory is ultimately constructive, not preservative, as researchers across disciplines have found. Their findings contradict early assumptions about memory's archival nature. Brockmeier (2015, 57), for instance, challenges AVM and its role in shaping memory studies, writing that "each act of remembering an experience is itself a new experience which, in the very act, subtly transforms the memory of the 'old' experience." During this transformation, he remarks, "we might add new emotional values, new beliefs, and even new knowledge to our memory […] fusing all of it with what we all the same consider an authentic and original memory" (57). The point here is that, although remembering may *feel* to the remembering subject like accessing an archive, it is in fact a process during which information and experiences are continually combined and recombined.

Alba and Hasher (1983) claim that this generative, constructive process of combination might happen at four different stages of encoding: selection, abstraction, interpretation, and integration. During selection, information is selected for encoding according to pre-existing schemas for remembering, and the information's relevance to the appropriate schema. In abstraction, the meaning of the selected information is abstracted from the syntactical and lexical context in which it

[4] Dummett frames his discussion in terms of knowledge. I take a broader view, thinking more in terms of content than knowledge. My account should still be compatible with Dummett's, though, since transmitting or preserving knowledge entails transmitting or preserving content.

is embedded; in other words, the message's meaning, but not its format, is coded. As Michaelian (2011c) points out, we can see in these first two stages that identifying content construction with generation is mistaken: both the first and second stages are constructive but involve content elimination, not generation. Following abstraction, interpretation involves accessing relevant prior knowledge and using it to form connections between the incoming stimulus and other knowledge. Finally, integration is the production of a coherent representation by blending together the products of the previous stages. As we can see, encoding is far more constructive a process than the image of a seal making an imprint into wax, as the wax tablet metaphor would lead us to believe! The wax tablet and the AVM it illustrates remain inadequate representations of how remembering unfolds.

While Alba and Hasher (1983) limit their discussion to construction during encoding, there are broader accounts that address the constructive nature of memory during retrieval, the later stage of remembering. Robins (2016a) categorises these accounts into three groups: connectionist, gist-based, and episodic hypothetical reasoning. The differences are subtle but important.

Connectionist accounts (e.g., Sutton 1998) are distinguished by their rejection of the AVM thesis that memory stores particular details of events. Instead, they claim that memory preserves patterns of semantically and conceptually linked ideas, which are grouped together such that the activation of one activates the whole set. For the connectionist, remembering a certain event is a result of forming associations based on general similarities from one experience to the next.

Gist-based accounts (e.g., Michaelian 2011c), by contrast, do not emphasise the similarities from one event to the next, but rather understand remembering as being a matter of forming general representations of single events. In other words, they capture the *gist* of what happened. While the precise details of an event are lost, those that are most likely to be relevant in the future are retained. The subject is

equipped with relevant knowledge for the future, while conserving cognitive resources.

Episodic hypothetical reasoning (EHR) accounts (e.g., De Brigard 2014; Michaelian 2016a) are distinguished by their rejection of the notion that memory is a distinct cognitive capacity. Proponents of EHR accounts point to the functional, phenomenal, and anatomical similarities between remembering and imagining counterfactual situations or planning for the future. Remembering, they claim, is not an act of preservation, but rather the deployment of a larger cognitive capacity that simulates situations to generate either plausible outcomes (De Brigard 2014) or accurate imaginings of the past (Michaelian 2016a); the temporal orientations of the products of these simulations —i.e. whether the simulation has a past, present, or future temporal valence—is what distinguishes remembering from the others (Robins 2016a).

Although connectionist, gist-based, and EHR accounts of memory offer different explanations of memory's architecture, all accounts converge on a central thesis: *that remembering is not, as supposed, the reproduction of a singular, discrete event from the past, but an act of reconstruction using information from a multitude of available sources* (Robins 2016). Memory science has repeatedly confirmed this point (e.g., Bartlett (1932); Brainerd and Reyna (2005); and Schacter, Norman, and Koutstaal (1998)). Because these findings and their corresponding constructivist accounts are incompatible with AVM, we have good reason to reject AVM and adopt a constructive model—or at least the central constructive thesis—in its place. I take it that the case for constructive memory has been well-defended in the literature (see e.g., De Brigard (2014); Michaelian (2016a; 2011c); Robins (2016b); Schacter and Addis (2007); Sutton (1998); and Sutton and Windhorst (2009)), so I have discussed it only briefly here, but I will add this last note: though different versions of constructivism offer different accounts of how memorial construction works, the important thing is that remembering is a process of continual combination and

recombination of information and experiences (Brockmeier 2015, 57). Importantly, *the resultant representations can contain more content than the original experiences that they apparently depict*.

At this point, it is clear that the transmissive analogy fails: if it is the case that memory is constructive and testimony is transmissive, they are fundamentally different. So, the analogy that so many philosophers have pointed to is simply incorrect, and we should abandon it altogether.

1.4. Testimony is constructive

Above, I showed that remembering is constructive, and that because this is the case, the analogy—as it has been described in the past—fails. However, if there are certain features of memory that compel us to think of memory as constructive, we would be remiss to ignore the fact that the same features are present in testimony. The transmissive analogy is misguided, but it will turn out that memory and testimony really are analogous; the analogy, however, will not rest on the transmissive features of memory and testimony. Instead, it will rest on their constructive features.

When I say that testimony has constructive features, what exactly do I have in mind? In other words, *how* is testimony constructive? Lackey (2008) argues that accounts of testimony focussing on the transmission of *beliefs* are incorrect, and we should recenter our discussion around *statements*. Crucially, she claims, recipients do not learn from testifiers' beliefs, but their words. It is this account that provides the grounds for restoring the analogy. In this section, I combine Lackey's account with Scott-Phillips' (2015) ostensive-inferential communication model, and address mindreading (Shanton and Goldman 2010) and its pivotal role in interpersonal communication (Nichols and Stich 2003).[5]

[5] There is a vast body of literature on these topics, spanning philosophy, psychology, and linguistics. Engaging with it in full would render the chapter nearly unwieldy. For the sake of simplicity, I rely on a few sources to paint a generic picture of the ways in which testimony might be constructive.

Most accounts of testimony, Lackey (2008) claims, are inadequate because they focus exclusively on either how testimony is an intentional act by a testifier, or how it is a source of belief for a recipient. Lackey's Disjunctive View of the Nature of Testimony seeks to remedy this situation by defining testimony as an act of communication *a* through which a testifier intends to convey information, or the content of which acts as a source of belief (Lackey 2008, 35–36). Since only one disjunct needs to be satisfied for *a* to qualify as testimony, we can analyse testimony as either intentional act or source of belief. Accordingly, we may consider the unique but complementary roles played by testifier and recipient, a task made easier by Lackey's supplementary definitions of speaker testimony (s-testimony) and hearer testimony (h-testimony): an act of s-testimony obtains when a testifier performs an act of communication through which she intends to convey information (Lackey 2008, 30), while an act of h-testimony obtains when a testifier performs an act of communication that a recipient reasonably understands as conveying information (Lackey 2008, 32).

There are two points to be made about the act of communication described in these definitions. The first is that the definition is broad enough to include gestures and writing alongside the more traditional category of literal speech. The second is that although s-testimony requires that the testifier "intend to express communicable content," it does not require that the testifier "intend to communicate *to others*" (Lackey 2008, 28, emphasis added). Thus, a given act of communication can constitute an act of testimony even if it was not so intended by the testifier, provided that the act of communication serves as a source of belief for the recipient.

Lackey's account affords us the opportunity to consider the roles played by testifier and recipient. Consider, as Scott-Phillips (2015) does, the character Hymie from the TV show *Get Smart*. Hymie is a humanoid robot with no ability to process idiomatic or metaphorical speech. Because he takes everything literally, in response to a speaker who says "Give me a hand!" he removes his own hand and passes it to

the speaker. As Scott-Phillips claims, this indicates that human communication involves a level of interpretation beyond simply literal phrases. For him, owing to the underdeterminacy of language, every single utterance can have an indefinite number of meanings. The underdeterminacy thesis states that

> "the linguistic semantics of the utterance, that is, the meaning encoded in the linguistic expressions used, the relatively stable meanings in a linguistic system, meanings which are widely shared across a community of users of the system, underdetermines the proposition expressed (what is said)" (Carston 2002, 19–20).

It may be tempting to think that here, Scott-Phillips is merely repeating the Gricean point that any given utterance bears both a sentence-meaning (the literal meaning of what is said) and a speaker-meaning (what the speaker intends to communicate) (Grice 1975; see also Turnbull 2003, chap. 4). But one crucial difference between Grice's and Scott-Phillips' accounts is that while Grice takes sentence-meaning to be fixed and speaker-meaning to be a matter of interpretation or inference, Scott-Phillips claims that even sentence-meaning needs to be determined by conversational participants (Scott-Phillips 2015, 18; see also Vicente and Martínez-Manrique 2008). Consider the phrase "Give me a hand!" Although it usually means "Help me," the phrase could have quite different meanings when uttered by a clockmaker (who needs the hands of a clock), a mannequin assembler (who needs to attach the mannequin's hands to the arms), or an attention-seeking violinist (who is demanding applause). Note that the first two are equally literal, yet have different meanings. Although Hymie gets "Give me a hand!" wrong, he must perform *some* interpretation, since communication is more than just a series of noises or symbols that magically produce a thought in an audience. As Scott-Phillips (2015) points out, if Hymie performed no interpretation *whatsoever*, it is not only idiomatic phrases that would confuse him—every phrase would. Exactly which meaning is intended by any act of communication will depend on a variety of factors, including, for example, the environmental context of the exchange and the participants'

conversational and personal history. While Grice's account fails to accommodate cases where the same sentence bears not only different speaker-meanings, but different sentence-meanings, Scott-Phillips' account can explain these cases by appealing to the role that interpretation plays in understanding.

The underdeterminacy thesis is, however, controversial. Semantic minimalists, notably Cappelen and Lepore (2005), claim that "a well-formed sentence does not need any contextual completion of the kind suggested by most pragmaticians in order to express a thought" (Vicente and Martínez-Manrique 2008, 394). If minimalism is right, then natural language can carry thought, and if natural language can carry thought, it is harder to identify the role of interpretation in testimony. But the cases for the underdeterminacy thesis and against minimalism are, to my mind, persuasively argued by Vicente and Martínez-Manrique (2008). I lack the space to rehearse the argument in full here, but importantly, they echo and provide compelling examples for Recanati's (2001) point that the presence of underdetermined terms is generalised: that an endless (or at least, unmanageably long) sequence of specifications is needed to disambiguate terms. Furthermore, on a minimalist view, sentences have complete truth conditions, and so a sentence such as "John is ready" is true iff John is ready. It seems, though, that we do not entertain such general thoughts: introspection suggests that thoughts are much more specific and complete than this, and furthermore, "if thoughts are to be responsible for our actions, they had better be more explicit than these putative minimal propositions. What would we do after tokening the very general thought that John is ready, punkt?" (Vicente and Martínez-Manrique 2008, 395). Though it may be the case that we occasionally entertain thoughts consisting of purely semantic, general information, such thoughts are atypical.

Still, it is not clear that we need to go quite as far as Scott-Phillips does when he says that language is *radically* underdetermined. In fact, we do not need to endorse a version of the underdeterminacy thesis *à la* Carston (2002) at all. We can

even defend minimalism *à la* Cappelen and Lepore (2005), and still get to where we are going. Not much hinges on the idea that language is underdetermined. What really matters for my argument that testimony is constructive is that there is some interpretive work done by testifier and recipient alike, and even if language is fully determinate after disambiguation, as the minimalists claim, there is a strong case to be made that the disambiguation itself will require interpretive labour from testifier and recipient. With this in mind, I offer a brief description of how interpretation plays a role in testimony.

Perhaps the most interesting determinant of interpretation is mindreading, which is "the capacity of ordinary people to understand the mind" (Nichols and Stich 2003, 2). During a communicative exchange, mindreading enables subjects to determine other subjects' background knowledge, motives, and intentions (Scott-Phillips 2015). Testifiers and recipients alike must cooperate to produce and interpret evidence of their own and each other's mental states and background knowledge (Sperber and Wilson 1986), and must provide evidence appropriate for that particular audience (Scott-Phillips 2015). When we couple the need for mindreading with the need for recipients to evaluate a given testifier's trustworthiness, sincerity, competence, motives, and intentions (Fricker 1994; 1995; Lackey 2015; Lehrer 2006), we learn that testimony is never a matter of producing a straightforward, literal utterance: in every situation, testifiers must determine an appropriate way to communicate an idea to a recipient, and must monitor the recipient continually to confirm understanding. In Scott-Phillips' words, both must establish "common ground: the information (the mental representations) that is known to two (or more) individuals and which both of them know, or believe, that the other knows" (Scott-Phillips 2015, 64). All of this happens in addition to the statement's literal meaning, which, it turns out, is only one input into a deeply layered and complex process; one that the transmissive model simply fails to capture.

The point here has been to show that interpretation is vital to communication. But a description of interpretation only tracks the recipient's cognitive task in receiving testimony, and thus tells only half the constructivist's story about testimony. To complete the account, we must also consider the testifier's role in constructing testimony. Here, it is worth emphasising the point that, just as it is the case that "construction" in memory is not synonymous with "generation" (Michaelian 2011c), neither should we identify testimonial construction with generation. The production of testimony may involve the elimination of some content as part of a larger constructive process. A testifier with a message to send needs to determine what the important features of the message are. That message itself is abstract and needs to be converted into an act of communication; the conversion requires selection of the abstract message's relevant details, which must then be organised according to linguistic rules and then communicated to a recipient. Developing the communication will require predictions and judgments about factors like what the recipient already knows, the recipient's motives, and the environmental context of the exchange. The testifier must depend on the recipient to do the same kind of mindreading. For example, if communicating sensitive information, a testifier who fears being overheard might lower her voice or be especially vague, counting on the recipient's ability to 'fill in the blanks.' She might also depend on the recipient's successful inference (from her lowered voice, for instance) that the information is sensitive, and subsequent actions to keep the information private. Since language is underdetermined and even simple sentences can have multiple meanings, both need to cooperate to establish mutual understanding and make testimony possible. In fact, we can even go so far as to say that testimony can generate content. In part, this content generation is a result of mindreading, but it could also be the case that testifiers and recipients contribute and rely on their own background knowledge to understand what is being communicated.

In testimonial exchanges, not only is it the case that participants occasionally mindread, make reparations, and 'fill in the blanks,' but that these features of communication are the norm rather than the exception. Scott-Phillips goes so far as to claim that we *always* engage in these activities. I am inclined to agree, but for the reader who finds that claim distastefully strong, it should be enough just to claim that these features play a central role. Thus, we see that just as memory involves construction, testimony involves ongoing interpretation. Construction and interpretation share important features, one of which is the practical benefit of reducing the amount of processing power required to remember or to communicate. Rather than storing a vast multitude of discrete memories that preserve specific details and particulars, the remembering subject is able to generate hypotheses about what might have happened in the past (De Brigard 2014). In testimony, participants conduct ongoing simulations of the other's mental state that allow us to enrich and expand on the content of their testimony, and reduce the amount of information required to be encoded into a message. Importantly, both construction and interpretation enable the introduction of new content, and thus lead to the surprising and counterintuitive conclusions that, despite Bernecker's (2008; 2010) and Goldberg's (2001) arguments against the generation of content by memory or testimony, that *one can remember more than one experienced, or understand more than a testifier communicated*.

This last claim should not be taken to imply that this means that all our memorially or testimonially based beliefs are false. I follow Sutton and Windhorst (2009) in resisting the idea that construction in memory inevitably results in distorted, inaccurate, or false memories. I claim the same about testimony: interpretation does not inevitably lead to distortion or misunderstanding. It is, rather, a process without which testimony would never occur. To say that testimony and memory are constructive is not to say that the contents are entirely fabricated, but to emphasise that both require the selection and integration of information from

a multitude of sources. The counterintuitive conclusion we thus reach is that content dissimilarity—in other words, when a recipient understands more than the testifier communicated—indicates that interpretive faculties are working correctly, not incorrectly.

There is a natural worry here that the possibility of understanding more than what was communicated risks the possibility that one can understand *any* proposition from *any* piece of testimony; that there need not be any connection between what is communicated and what is understood. We need not be alarmed, though. In memory, philosophers try to identify conditions for what constitutes a genuine memory, which has downstream consequences for a subject's claim to know that *p* on the basis of her memory. So should it be possible to set down limits on the type and strength of the connection between what is communicated and what is understood. Just as philosophers need to establish some conditions differentiating successful from unsuccessful remembering, so too do they need to establish some conditions that work to restrict how much, what kind, and by what process new content can be introduced into testimonial exchanges. There have been forays into this project: Lackey (2008), for one, has stipulated that a "reasonably obvious connection" must exist between "the content of the proffered act of communication and the content of the proposition testified to" (30). It should be noted here, though, that to say that there needs to be a connection is to say that the content of the testimony must have some kind of match with the resultant representation. Requiring that there be such a match does not presuppose that one endorses a causal theory of either memory or testimony (i.e. the connection in question is about the content, not the causal link between what is testified and what is received). It is not my project here to determine what constitutes a "reasonably obvious connection," nor indeed what any of the other limits of accuracy are, but it *is* my project to show that such considerations need to be made in both memorial and testimonial cases.

One final note should be made here before moving on: I argued above that noise could contribute information to the intended message in testimonial exchanges. It may seem that there is no memorial analogue for noise, but I think this is mistaken: it is easy to imagine signals that might compete with what one appears to remember. For example, alcohol consumption (Lister, Eckardt, and Weingartner 1987), sleep deprivation (Stickgold 2013), exposure to distinct events that are sufficiently similar that one collapses them into a single representation (Devitt et al. 2016), or exposure to a conflicting account of an event that one has experienced (Loftus 1997b) can all influence what is recalled; indeed, each might contribute information to the message that is understood.[6] It does seem, then, that noise-like phenomena appear during ordinary remembering as they do during testimonial exchanges.

1.5. Drawing an analogy between memory and testimony

It is not difficult to understand why figures like Dummett endorse an analogy between memory and testimony: intuitively, both transmit but neither generate content. This version of the analogy, however, is inadequate in that they are grounded in faulty understandings of memory and testimony. It is a mistake to characterise memory as (strictly) preservative and testimony as (strictly) transmissive, and therefore the transmissive analogy fails. But if we abandon, as we are right to do, the preservative and transmissive views of memory and testimony that underpin the transmissive analogy, and embrace constructivist accounts instead, it will turn out that the two really are analogous after all.

It may seem that endorsing a certain version of constructivism about memory (detailed above in Section 1.3) would bear on the plausibility of the analogy, depending on whether testimonial construction is sufficiently similar to memorial

[6] Justine Kingsbury rightly pointed out to me that alcohol consumption and sleep deprivation might better be understood as factors that increase the likelihood of noise effects, but not as competing signals themselves.

construction. While the account of how memorial construction works will depend on the specific version of constructivism one endorses, it should suffice for my purposes to adopt a generic version of constructivism. This is because the central thesis I advance about testimony—namely, that testimony is constructive in virtue of the integration of information from multiple sources—is broadly similar to the central thesis about constructivism in memory, and the specific details and variations on the central thesis do not threaten my argument.

Because both remembering and testifying involve simulation and interpretation, a subject can remember something more than what was experienced, or understand something more than what was said. The similarities between memory and testimony as they function as sources[7] of knowledge cannot be ignored; importantly, they have parallel epistemic and epistemological implications, including, most crucially, the ability for a receiver to gain more than what is expressed by a source. Thus, if this analogy holds, it creates room to apply the same debates from epistemology of testimony to epistemology of memory (and vice versa). But the potential products of this analogy do not end there. Not only is the analogy of epistemological significance, it would also allow us to conceptualise memory and testimony in similar terms, to investigate their metaphysical similarities and differences, and to explore the ethical questions that arise in one while shifting our focus to the other (for instance, concepts such as Miranda Fricker's (2007) testimonial injustice, which I take up seriously in Chapter 4, or Sue Campbell's (2003) relational remembering). The analogy is conceptually rich and has tremendous potential to bear philosophical fruit.

[7] I use the term "source" quite loosely. To be sure, there are lively debates over whether either memory or testimony can function as sources in any genuine sense, i.e. generate new knowledge (see, e.g., Graham (2006) and the exchange between Lackey (2005; 2007) and Senor (2007)). However, when I use "source," I am not weighing in on those debates. Those debates lie downstream of my claim, which is merely that memory's capacity to act as a (genuine or pseudo-) source of knowledge is coextensive with testimony's.

Although the features of memory and testimony indicate that there really is a constructive analogy between memory and testimony, and although endorsing the analogy might be tempting because of its potential implications, there are also reasons to resist it. There is, for instance, one disanalogy that is obvious right at the outset: although both memory and testimony act as sources of knowledge, they act as sources of knowledge about fundamentally different things. To a first (controversial) approximation, memory is a source of a knowledge about experiences in the past, and testimony is a source of knowledge about a state of affairs in the world, in propositional form. It might seem that the difference is irreconcilable and the project of drawing any analogy whatsoever is severely misguided, but I think this is mistaken. First, it may turn out that memory and testimony are both propositional (i.e. all features of episodic memory may be expressible as propositions) (Fernández 2006). If this is the case, then the analogy is clear. But even if episodic memory is not reducible to propositions—and the claim that it is reducible is controversial (Sant'Anna 2018b)—then we can still draw an analogy between the two. The important thing is that both carry content, and the content that they carry undergoes the same kind of transformation such that the processes, though they involve different components, are deeply similar. We must accept this disanalogy, but if we do, we will see that memory and testimony are otherwise broadly analogous. To show that they are indeed analogous, I next consider and respond to two worries about where the analogy's limits lie.

Before proceeding to these worries, it is worth addressing the following objection. The characterisation of testimony upon which I have relied throughout this chapter is that testimony is "the [testifier's] *intentional* expression of communicable content" (Lackey 2008, 28, emphasis added). But remembering, everyday experience suggests, can be done intentionally (by encoding deliberately) or non-intentionally (by encoding subconsciously). Because there are instances of remembering that lack an intentional component, while testifying is intentional, the

constructive analogy appears to fail. I readily concede that both remembering and testifying can be done intentionally, but I deny that testifying is only ever intentional in the sense that memory is intentional. In other words, I argue that both remembering and testifying can be done non-intentionally, and furthermore, that this is entirely consonant with Lackey's account of testimony.

Many of our experiences of remembering are non-intentional: we often navigate our lives without forming intentions to memorise information now for use in the future, and memories often surface without a subject's intention to retrieve them (Berntsen 2010). Both encoding and retrieval are processes that routinely unfold in the background, with little or no real direction from the subject. That we find ourselves reminiscing without meaning to do so, or recalling information we were not consciously aware we could access, indicates that the relevant memories were formed without our awareness or intention.[8]

A source's role in remembering non-intentionally is at the point of encoding at the time of the original experience. It is only in exceptional cases that we consciously encode with the express intention of making the information available for recall in the future. Yet, despite this lack of conscious remembering for the future, we still find ourselves able to recall events from the past. Similarly, a testimonial recipient can gain beliefs from a testifier, even if that testifier did not intend to instil those beliefs in the recipient. Initially, it seems odd to claim that one can testify non-intentionally, but I show next that the claim is supported by our everyday experiences of testifying.

To understand how non-intentional testifying is consistent with Lackey's characterisation of testimony as requiring an act of communication on the part of the testifier, recall that Lackey does not consider an act of communication to "require

[8] It is important to note that to claim that some instances of remembering are non-intentional is not to claim that remembering happens *despite* one's intentions. It is to claim that remembering might happen absent a subject's intention to do so.

that the speaker *intend to communicate to* others; instead, it requires merely that the speaker *intend to express communicable content*" (Lackey 2008, 28). The testifier might intentionally express communicable content without intending to express anything to anyone, but may nonetheless do so; if this is the case, the instance still qualifies as testimony. In other words, the *expression of content* is intentional, but the *testifying* is not. This is consistent with Lackey's h-testimony: a testifier's act of communication that a recipient reasonably understands as conveying information (Lackey 2008, 32). Separating the intention to express communicable content from the intention to testify allows for cases such as posthumous publications of private diaries to qualify as testimony, since the expression of communicable content—namely, the act of writing—was intentional, even if the author never intended for the content of the diary to be accessed by others. It is precisely because Lackey does not view the higher-level intention as being a necessary condition for testimony that the notion of non-intentional testimony is even coherent.

Recall from Section 1.4 that if a subject performs an act of communication intending to convey certain information, then S s-testifies that *p*, and that a subject performs an act of communication that serves as a source of belief for a recipient, then S h-testifies that *p* (Lackey 2008, 32). Unsurprisingly, in many cases, s-testimony and h-testimony overlap: both obtain where S testifies that *p* by performing an act of communication through which she intends to convey the information that *p*, *and* that act of communication also acts as a source of belief for the recipient. For example, if someone asks me what time it is, and I intend to instil the belief that it is 7:35 p.m. by saying "It's 7:35 p.m.," and my statement acts as a source of belief for the recipient, then both s-testimony and h-testimony have come to pass. But where h-testimony instantiates without s-testimony, there is non-intentional testimony. It is testimony in virtue of being h-testimony, and it is non-intentional in virtue of the fact that the testifier did not intend to communicate to others. Let us clarify this with an example. Imagine Isadora, a student who uses a calendar to manage her time. None of the

information is secret and she makes no effort to keep the contents of the calendar private. She records all her appointments, assignments, and classes in the calendar, and regularly refers to it. One evening, Isadora leaves her calendar on the kitchen table, where her brother picks it up and learns that Isadora has a biology midterm in the morning and a poetry assignment due the next day.

In this example, Isadora intentionally expresses communicable content by writing, and so performs an act of communication that is transformed into an instance of testimony once the calendar acts as a source of belief for Isadora's brother. In doing so, the act of communication satisfies the second disjunct of Lackey's view on the nature of testimony, making this case an instance of h-testimony. Such a case illustrates how one might testify without having any intention to do so.

The objection at issue here is that although remembering can be done intentionally or non-intentionally, testifying can only be done intentionally. This seems to be especially worrisome if we take seriously Lackey's claim that testifying involves intentional expression of communicable content: if testimony *qua* epistemic source necessarily involves intention, does it not threaten the analogy that memory seems to lack the intention required by this definition of testimony? I think not. By dissociating the intention to express communicable content from the intention to testify, as Lackey does, I have shown this objection to be mistaken. Both memory and testimony can be either intentional or non-intentional in the relevant sense; that is, one can remember without intending to remember, and one can testify without intending to testify.

1.6. The mindreading problem

I claimed in Section 1.4 that mindreading—the capacity of minds to represent other minds—is an integral feature of communication, and thus, of testimony. Testifiers and recipients need to form representations regarding each other's mental states,

motivations, background knowledge, and credibility. In testimony, I have argued, mindreading is a key component of interpretation, and is what makes testimony constructive. Mindreading allows participants to make reparations for unfinished sentences and misused words, to 'fill in the blanks,' and to discern why a given testifier is offering testimony or what a given recipient needs to know; in other words, it contributes content to testimony and thus plays a vital epistemic role.

The reader might wonder why we should worry about mindreading in particular. It might be unclear why, of all the features of memory and testimony, mindreading is the one that should concern us. Why would a disanalogy in this respect threaten the analogy at large? The answer is that this project is about the epistemic functions of memory and testimony. In particular, I am interested in whether memory and testimony provide knowledge in similar ways. Therefore, any aspect of either memory or testimony that influences its epistemic function—including transmission of information, generation of content, etc.—is of interest to my project here. And so, because mindreading contributes content to testimony, we are confronted with this question: is there mindreading in remembering? Without a plausible cognate, the analogy appears to fail. Furthermore, translating testimonial mindreading into memorial terms would come quite close to claiming that a subject mindreads her future self when committing something to memory, or mindreads her past self when recalling. These phrases have an odd ring to them, but I aim to show in this section that such mindreading is possible, and so the objection does not threaten the memory–testimony analogy.

A central question regarding mindreading is precisely how a subject represents in her own mind the mental states of another. As Shanton and Goldman (2010) characterise it, the simulation theory of mindreading holds that to attribute mental states to others, subjects simulate the mental processes of others' minds. To do so, subjects 'put themselves in the other's shoes' to determine what might be unfolding in that person's mind. A subject creates pretend mental states in her own

mind that correspond to the mental states of a target and uses those pretend mental states as inputs into a simulation. The simulation generates an output mental state, which the subject then attributes to the target.

What is particularly interesting about Shanton and Goldman's (2010) account is that they are not simply concerned with how subjects attribute mental states to others (interpersonal mindreading); they also offer an account of how subjects attribute mental states to themselves (intrapersonal mindreading). They claim that, given the mounting evidence that the same (or a very similar) cognitive mechanism involved in simulation is deployed in a broad range of cognitive processes, including mental time travel (MTT), or projecting oneself into the past or future.[9] The parallels between simulation in MTT and in mindreading are striking (Shanton and Goldman 2010), and have not gone unnoticed by psychologists and neuroscientists (e.g., Buckner and Carroll 2007; Hassabis and Maguire 2007; Spreng, Mar, and Kim 2008). Even earlier than these writers, Gordon (1986) argues that predicting one's own future behaviour and predicting others' behaviour and attributing certain beliefs to them is a function of the same capacity for simulation. MTT and mindreading are similar in at least two respects: first, in MTT, the mental time traveller sheds present mental states and project herself into the past or future, just as the mindreader sheds her own mental states to adopt those of a target. Second, both simulational mindreaders and mental time travelers construct their target states through the combination of details from the past (Shanton and Goldman 2010). The upshot of these similarities is that the cognitive mechanism that is deployed to enable simulations in mindreading is also deployed during past-

[9] The cognitive mechanism at work here is likely also deployed during other cognitive tasks, such as imagination and navigation (Mahr and Csibra 2018), but that discussion is beyond the scope of this chapter.

and future-oriented MTT (Buckner and Carroll 2007; Shanton and Goldman 2010).[10] Though Shanton and Goldman come quite close to arguing that remembering is a kind of mindreading (Michaelian 2016a), I am not committed to so strong a view; I claim, more modestly, that remembering includes activities that are very much like mindreading.

My aim is to show how mindreading could play a role in remembering, even though there is only one subject. In the ordinary case of interpersonal mindreading, a subject must manufacture pretend beliefs based on factors such as circumstances, competence, and desires or motives affecting the target, and generate an output based on those pretend beliefs. To have a chance at doing so successfully, the subject must 'step into another's shoes' and construct or simulate what it *would be* like to be that person, to generate a hypothesis of what another could be thinking. In memory, things are somewhat different: because the subject is remembering her own past, she already knows what it was like to be herself at that past time. And so, in this non-paradigmatic intrapersonal case, the subject does not simulate what it *would be* like, but what it *was* like to be that person.

A few examples will help to clarify how the ordinary case of mindreading can be adapted from interpersonal to intrapersonal. What is important here is that the subject is trying to understand things from his past self's perspective. Consider a situation in which the subject retrieves a memory of event *e*. When he projects himself into the past to try to understand things from his past self's perspective, he

[10] One problem for this view is the evidence that subjects with medial-temporal lobe amnesia can pass false-belief tests (Rosenbaum et al. 2007). We should expect difficulties in mindreading tasks from individuals with MTT deficits, so the fact that these individuals can pass false-belief tests—paradigmatic measures of mindreading—suggests that mindreading and MTT are not subserved by the same cognitive mechanism(s). Though it may not be the case that the cognitive mechanisms that subserve mindreading and MTT are perfectly coextensive, it should be enough for my argument that there is a significant overlap. Even if they are different at the neurological level, MTT and mindreading seem to play similar roles in learning from memory and learning from testimony, respectively. After all, for there to be an analogy between memory and testimony does not require that the two be identical in every respect; it merely requires that the two play analogous roles. This seems to be the case even if it turns out that MTT and mindreading are not subserved by the same cognitive mechanism(s).

might recall that he was angry at the time. He might remember his own accelerated heart rate, his proverbial tunnel vision, and might realise that in that moment, he was not very observant of the world around him. Alternatively, he might remember that *e* happened late at night. He may feel the fuzziness of his own mind, the heaviness of his eyelids, the slower speed at which the world seemed to pass by in that moment. He might feel his own past self's difficulty in observing and encoding. He can relive his own mental state, much like testimonial exchange participants try to understand each other's mental states. In either case, knowing that his memories are spottier when he was tired or angry at the time the event occurred may cause him to cast moderate suspicion on what he seems to remember, and be cautious in taking a given memory as conveying the information that *p*. Similarly, certain judgments about the circumstances under which a testifier offers testimony may also influence whether and how that testimony is taken up.

A subject can also mindread her past self's desires and motivations. Imagine a teenager trying to explain a poor decision to her parents by saying "It seemed like a good idea at the time." She might know that she would not make the same decision again, but she can put herself in her past self's shoes, and understand that her past self believed that decision to be a good one. It should be noted, however, that it is not necessary for the subject to defend their motivations aloud for us to say that mindreading has taken place. As long as the subject attributes desires to the past self, there has been mindreading.

So far, I have shown how mindreading plays a role at the point of memory retrieval; that is, how a subject might mindread his past self. This alone, however, does not establish that mindreading plays the same role in memory that it does in testimony. After all, I have only shown that receiver mindreads source, but mindreading is a two-way relation in testimony. To show that mindreading plays the same role in memory and testimony, we would need to find a case of mindreading in which source mindreads receiver; in the testimony case, this would

be a testifier mindreading his recipient, but in the memory case, it would involve the earlier self mindreading her later self when attempting to make information available for later recall.

Let us consider this example: Say a subject, Hector, knows he has a long day of travel soon. Because he has travelled before, he knows that at the end of the trip, he will be tired and mentally scattered. He can imagine things from his future self's perspective, and predict that he will have difficulty recalling things like the name of his hotel or the train route he needs to take to arrive there. Although Hector might be able to remember these details under normal circumstances, he can also imagine that his future self, having been in transit for a long time, will be less able to recall them.

In this example, because Hector can imagine things from his future self's perspective, he is able to take steps to minimise the impact the travel has on his ability to function. He might, for instance, compile a detailed itinerary that he can refer to upon arriving at his destination. Hector's situation shows that it is indeed possible for a subject to predict how future circumstances might affect his cognitive capacities, much like a testifier can predict how a recipient's circumstances might affect how he receives testimony.

It is also possible for a subject to predict her competence in different domains. Imagine, for instance, that I reliably remember birthdays and appointments, but have a very poor memory for lists of items. Knowing this fact about myself, I might rely only minimally on a calendar to keep track of dates; I can imagine that my future self will be able to recall this information without difficulty. However, if I can imagine entering a grocery store and suddenly being distracted by the available products, I might not trust myself to remember what groceries I need to buy without writing them down. In doing so, I am imagining what it will be like to be my future self, and adjusting my behaviour based (in part) on the type of information to be retrieved. We can see here, then, that competence is domain-

specific, and that mindreading with respect to competence is possible in the memory case just as it is possible in the testimony case.

We may worry that predicting facts about one's future self and her mental states is an unreliable enterprise: even under the best circumstances, predictions about the future are often wrong. One might therefore object and claim that we cannot mindread reliably in remembering contexts. At worst, however, this worry only points out a problem with the effectiveness or reliability of mindreading, not the presence of it. Indeed, it is fair to point out that we are not very good at mindreading in the case of memory; a prime example would be believing that one's memory is infallible and being unwilling to accept evidence contradicting what one seems to remember. While we might not be very successful at mindreading in the memory case, however, it is worth noting that we do not mindread perfectly in the testimony case either; any number of considerations a subject makes about her target could be mistaken. (We would expect, for example, that if we were strong mindreaders, we would see higher rates of deception detection, but in fact we detect lies at rates barely above chance (Bond and DePaulo 2006).) Thus, the fact that we might be mistaken in future-oriented mindreading does not threaten the analogy.

Above, I argued that one role of mindreading in testimony is to compensate for the underdeterminacy of language. Because testimony is typically presented in propositional statements, it is easy to see how underdetermination might present: simply put, a certain statement or sentence could have more than one meaning, and so could be expressing different propositions depending on what the meaning of the statement is. But memories are not clearly propositional, so it is more difficult to see how there might be a memorial analogue for testimonial underdetermination. In other words, because the medium of memory is not propositions, but episodes (*pace* Fernández 2006), the notion of underdetermination that applies to testimony does not seem to apply to memory. But if we are more open-minded about how

underdetermination might appear in non-propositional contexts, we can see how underdetermination plays a role in episodic remembering.

Episodic memories typically present themselves to us as though they come from particular past events. For instance, I might remember a cycling trip I took between two cities. If I have only done this one time, I can be fairly confident that I really am remembering a particular event. This, however, is not always the case: for instance, when I think of cycling to my office, I take myself to be remembering a particular trip (for example, yesterday morning's ride). But in fact, I have cycled to my office hundreds of time, so unless something very unusual and memorable happened yesterday, I have no obvious way to differentiate this apparent memory of yesterday's trip from any other ordinary ride. If I am to gain knowledge from this apparent memory, I need to perform some kind of inference to determine what the memory really represents and what kind of knowledge I can claim to have in virtue of what I seem to remember.[11]

If it sounds familiar to assert that a subject needs to perform an inference to determine what is really being conveyed by a message, it should: it is exactly what I have claimed above is at work in testimonial exchanges. Just as the underdeterminacy of language means that a given utterance could express different content, a given memory representation could carry several potential meanings. In both cases, subjects must use other information to infer what the message is really conveying.

1.7. The possibility of source-receiver interaction

I now turn to one final objection, which hinges on this *prima facie* difference between memory and testimony: in testimony, a recipient has the opportunity to ask a testifier for more information. By contrast, it appears that when remembering, a subject (the later self) has only limited information. Given the direction of time, it is

[11] I owe this view to Denis Perrin.

impossible for her to obtain information beyond what her earlier self has encoded. Her knowledge is restricted to the contents of her memory. Consequently, there is a difference between the limits of testimonial knowledge and memorial knowledge. Testimonial knowledge is such that it can be clarified, amplified, and updated almost immediately through conversation with a testifier, but memorial knowledge exhibits no such property. A testifier and recipient can engage in a verbal volley of a kind that is unavailable to a remembering subject: recipients are able to ask questions and update their beliefs in ways that rememberers cannot. Put simply, recipients of testimony have much more access to supplementary information than rememberers do, since rememberers cannot communicate with their past selves to gain more information. The possibility of updating is closed off once the original memory is formed. Thus, the objection goes, testimony acts as a much richer source of knowledge than memory does, and the ability of receivers to seek further information in one case but not the other poses a threat to the epistemological analogy in terms of the nature and accessibility of knowledge that is available through each source.

A few examples may help to underscore the difference between memory and testimony in terms of the availability of further information. First, imagine Beatrice, who is at work and realises she does not have her wallet. Upon this realisation, she remembers what she did this morning, and remembers that although she had her wallet as she was preparing to leave for work, she set it down on the kitchen counter to take one last sip of her morning coffee. She subsequently forms the belief that her wallet is on the kitchen counter. In this example, Beatrice forms a belief based on her memory. But what if she cannot form such a belief? What if she cannot remember what she did with her wallet?

Suppose that Beatrice has realised that she does not have her wallet, but cannot remember what she has done with it. She phones her partner Bertrand to ask whether he has seen it. He tells her that her wallet is sitting on the kitchen counter.

Beatrice subsequently forms the belief that her wallet is on the kitchen counter. In this modified example, as in the original above, Beatrice forms the belief that her wallet is on the kitchen counter. In this example, however, Beatrice can ask for more information: she can, for instance, ask whether her credit card is still inside, or whether she has any cash (and how much), or whether she left her office ID card next to her wallet. When receiving testimony from Bertrand, she is in a better position to seek new information than she is when relying on her memory alone. It seems, then, that the potential for gaining knowledge through testimony is higher than in memory.

Upon closer inspection, however, it becomes clear that the ability for that testimonial knowledge to be instantly updated is more a feature of conversation than it is of testimony *simpliciter*. When a recipient seeks further information by asking questions, she is still engaged in an epistemic project, but that project is not strictly testimonial. While we might readily agree that asking questions is both a component and type of conversation, it is less clear that asking questions should be considered part of testimony. Another modification to the example will help to illustrate this point. Suppose that just as Beatrice realises that her wallet is missing, she receives a text message from Bertrand that says, "You left your wallet on the kitchen counter." Beatrice subsequently forms the belief that her wallet is on the kitchen counter. She sends a reply asking whether her credit card is still inside, but receives no reply. Unbeknownst to Beatrice, Bertrand's phone battery has just died and he neither receives nor replies to her texts. In this final example, the difference between testimony and the broader category of conversation is highlighted. In this case, Beatrice attempts to gain further information from Bertrand, but because his battery has died, she is unsuccessful. But her inability to gain further information from her partner (and form corresponding beliefs) does not compel us to disregard the initial text message as an instance of testimony; rather, it emphasises the point that asking questions and gaining answers are not necessary features of testimony. Instead,

these kinds of exchanges are examples of conversation, broadly construed. When recipients engage in dialogue with testifiers, they are conversing, but not every utterance qualifies as testimony. Thus, there is no true disanalogy between memory and testimony here: although it is the case that testifiers and recipients *can* converse with each other in a way that rememberers do not converse with their earlier selves, what is paramount here is that *they do not always do so*, and yet the absence of back-and-forth communication does not disqualify these utterances as instances of testimony. Although the objection raised here might be an objection to establishing an analogy between memory and communication broadly construed, it poses no real threat to the narrower memory–testimony analogy I propose here. When we limit our analysis to testimony proper, excluding back-and-forth supplementary conversation, we find no disanalogy with memory.

1.8. Conclusion

Here, it bears repeating what I said at the beginning of this chapter: it is tempting to believe that there is some deep analogy between memory and testimony. The analogy aligns with our intuitions about how memory and testimony function, while also offering fertile ground for developing epistemologies of both. But since neither testimony nor memory are preservative or transmissive in nature, the transmissive analogy does not hold. But once we acknowledge that both memory and testimony are not transmissive but constructive, the analogy re-emerges. It has been my project in this chapter to reconceptualise testimony and memory along exactly these lines, and to show that the two really are analogous. Dummett, it turns out, was right all along, but for the wrong reasons.

In the epistemologies of memory and testimony, there are many lively debates too broad and diverse to capture here. The aim of this chapter has not been to address these debates, nor to posit any solutions; those controversies lie downstream of my project. But if testimony is constructive, and we are right to think

that there is a constructive analogy between memory and testimony, we should expect that the questions in one will translate to the other. So too, I imagine, will the answers. The analogy I have developed in this chapter grants us all the advantages of the transmissive analogy—for instance, allowing us to apply epistemology of one to epistemology of the other—while maintaining consistency with empirical research. It has been my aim to show that memory and testimony are sufficiently similar that applications of issues in the philosophy of memory controversies to philosophy of testimony, and vice versa, are both possible in the first place. It is very exciting indeed to find that the memory–testimony analogy, long-persistent in philosophers' minds, can find real purchase.[12]

[12] I am grateful to audiences at the 2016 New Zealand Association of Philosophers' conference, the 2017 Issues in the Philosophy of Memory conference, the University of Otago philosophy postgraduate seminar, and the 2017 Australasian Association of Philosophy conference for feedback that greatly contributed to the development of this chapter. I am also grateful to Zach Swindlehurst for proofreading.

CHAPTER 2:
TESTIMONY-HOW

2.1. Introduction

Thinking about the relationship between memory and testimony in Chapter 1 invites us to wonder about how far the analogy between the two might go. One point at which it might appear to break down, for instance, can be found at the taxonomic division between declarative and non-declarative memory. While my argument in Chapter 1 established an analogy between testimony and declarative memory, we do not so far have any principled reason to think that there is a further analogy between non-declarative memory and testimony. Although I do not offer an extended argument in favour of this extension of the analogy, in this chapter I aim to address that concern by providing a suggestion as to how we might argue for such an extension. The very name "non-declarative memory" hints at the obvious objection one might have to extending the analogy: by definition, non-declarative memory is *non-propositional*, but *testimony* is propositional, and it is much harder to understand an epistemological analogy between two sources of memory that differ on such a fundamental level. This chapter can be understood as an extended response to this concern; for the duration of it, I bracket memory altogether and focus only on testimony.

Like Franklin the turtle from the children's book series, I know how to count by twos and tie my shoes. I can zip zippers and button buttons. I can also ride a bike, knit, speak French, change a flat tire, type, and operate a lawn mower. How do I know how to do these things? Because someone taught me how to do them.

Since Coady's (1992) landmark study on testimony, several different accounts of testimony have emerged. Each responds to supposed problems of its predecessors, and the differences between the accounts are sometimes minor and other times quite pronounced. Nevertheless, the accounts converge on the basic

point that testimony comprises assertions or utterances offering information about a state of affairs. In other words, received wisdom has it that propositions—truth-apt statements about the world—are the currency of testimony.

However, the way that we deliver instructions is often in the imperative mood, such as "Turn left and go down the stairs," "loop the yarn counter-clockwise around the knitting needle," or "lather, rinse, repeat." None of these instructions are strictly propositional; they cannot be true or false any more than it can be true or false that "Shut the door!" Yet, instructions seem to instil some kind of knowledge (albeit incomplete knowledge) in instructees: how to get to a certain room, how to perform part of a knit stitch, and how to shampoo one's hair. The knowledge gained is not perceptual, not introspective, not memorial, and not rational, nor is it synthesised from a combination of these sources. Coming to know how to knit, for example, depends crucially on an instructor's knowledge of how to knit and her delivery of that knowledge to an instructee. If that is not testimonial, then it is hard to see what else it might be.

This conundrum may be presented as an inconsistent triad of commitments:

i. All instructions are a kind of testimony.
ii. Some instructions are non-propositional.
iii. All testimony is propositional.

If it is true that testimony deals exclusively in propositions, then instructions ought not to count as testimony in virtue of the fact that they are non-propositional. Yet, instructions ought to qualify as testimonial. So, the solution is to deny (at least) one of (i), (ii), or (iii). In this chapter, I consider what commitments each solution would involve, and whether each solution would indeed be tenable as an answer to the triad. Ultimately, I show that rejecting (i) and (ii) are not feasible strategies, but rejecting (iii) provides us with a viable way forward when it comes to understanding exactly how information about how to do something can travel from one agent to another.

This chapter is structured as follows: In Section 2.2, I discuss some preliminary considerations about knowledge-how and giving instructions. I proceed with a detailed discussion and defense of commitments (i) and (ii) in Sections 2.3 and 2.4, respectively. In Section 2.5, I argue that (iii) is the least tenable of the commitments, and argue that we ought to reject it in order to resolve the inconsistent triad. If it is correct to deny that testimony is necessarily propositional, then such a denial might seem to mean that much of contemporary epistemology is wrong. I think this response would be going too far. We do not need to restart the epistemology of testimony from scratch; rather, we need to supplement existing accounts of testimony. The final sections of this chapter begin this task. In Section 2.6, I argue that giving instructions qualifies as a specific kind of testimony, namely, testimony-*how*. Section 2.7 introduces a basic account of testimony-how, including discussion of how testimony-how compares to paradigm cases of testimony. Finally, in Section 2.8, I consider the analogy between memory and testimony and how it looks once we consider that testimony is not necessarily propositional.

Let us begin with some preliminary remarks.

2.2. Preliminary remarks

I will begin by addressing a possible objection to my entire project. I could be accused of advancing a circular argument of the form that testimony isn't propositional because instructions are testimony, and instructions are testimony because testimony isn't propositional. I understand where such an objection might come from: after all, it might look a bit like I am simply cutting the Gordian knot by amending the definition of testimony to suit my purposes. But I think the objection is misguided. My interest here is about the *nature* of testimony, not its *definition*: our work on testimony has been shaped by an assumption that the thing of interest when we talk about testimony is propositions. This assumption has, of course, informed our definitions of testimony. But, as I will go on to show, we would do

well to abandon the assumption. This will, of course, influence the downstream definitions of testimony, but is most fundamentally about the nature of testimony: whether testimony really does trade exclusively in propositions, or whether propositions are just one currency in the economy of testimony.

2.2.1 Intellectualism and anti-intellectualism

It seems fairly clear that one agent can bring another to believe (and possibly know) that *p* on the basis of her testimony. Knowledge-that is shared among agents through testimony; on this, epistemologists basically agree.[1] My project here is to explore how agents might bring each other to know *how* to do something.

An important thing to take away from this section is that the status of knowledge is, to some degree, independent of what testimony is. What matters is that we understand what the currency of testimony-how ought to be: namely, an agent's coming to know *how* to φ on the basis of testimony. We need to understand what knowledge-how really *is* before we can begin to understand its relationship to testimony.

The relationship between testimony and knowledge is not the only relationship of interest: epistemologists have also sought to understand the relationship between knowledge-that and knowledge-how. *Intellectualism* and *anti-intellectualism* are opposed positions in the debate. Intellectualists hold that knowledge-how is fundamentally knowledge-that: that procedural knowledge is just a species of propositional knowledge. On this view, knowing how to perform an

[1] There is a lively debate amongst epistemologists about whether knowledge-that is appropriately characterised as *shared* through testimony, and if so, in what sense—i.e. whether beliefs are duplicated, transmitted, or generated anew—as well as the epistemic status of the recipient's belief and the conditions under which it rises to the level of genuine knowledge. I bracket this debate for now, since there is nevertheless general consensus that one agent can bring another to believe that *p* on the basis of her word. The exact details of how this comes to pass are still under debate, but for my purposes, don't matter. What matters for me is that whatever is going on at the epistemological level for knowledge-that, it seems that there should be some analogous way that agents share knowledge-how.

action requires some prior knowledge of rules or facts that guide the action; in other words, knowledge-how depends on some way on knowledge-that. Anti-intellectualists deny this thesis, either by asserting the negation or by asserting the opposite, i.e. *weak anti-intellectualists* deny that knowledge-how depends on knowledge-that, while *strong anti-intellectualists* invert the hierarchy and hold that knowledge-that in fact depends on knowledge-how (Fantl 2008).

Most epistemologists accept that knowledge-how and knowledge-that are distinct kinds, at least on the surface. The real disagreement about knowledge-how and knowledge-that arises when we try to cash out the relationship between the two. Those who hold that knowledge-how reduces to knowledge-that—i.e. that knowledge-how is a species of knowledge-that—are *intellectualists*. This view is held, notably, by Stanley and Williamson (2001), but has a long lineage of defenders including Bengson and Moffat (2011a), Brown (1971; 1974), Ginet (1975), Katzoff (1984), Snowdon (2003), and Vendler (1972).

Ryle (1949, chap. 2) is often credited with catalysing the intellectualist/anti-intellectualist debate with his rejection of what he called "the intellectualist legend" (Bengson and Moffett 2011b) and defense of what Fantl (2008) calls *weak anti-intellectualism*. On Ryle's view, knowledge-how does not depend on knowledge-that, and knowing the right facts about how to φ is neither necessary nor sufficient for knowing how to φ. Since Ryle, many others have defended weak anti-intellectualism, including *inter alia* Carr (1979), Lewis (1990), Noë (2005), and Riley (2017).

An even stronger version of the anti-intellectualist thesis is what Fantl (2008) calls *strong anti-intellectualism*, which is the view that knowledge-that depends on knowledge-how (Hartland-Swann 1956; 1957; Hetherington 2006; Roland 1958). Habgood-Coote (2018a) and Williams (2008) offer intermediate or hybrid views. I bracket these strong anti-intellectualist and intermediate views for this chapter.

The point of this section has been to emphasise that the category of procedural knowledge is wildly variable both with respect to its relationship with propositional knowledge and the degree to which procedural knowledge rests more on algorithm and learning a set of propositions (e.g., Bengson and Moffett 2011b) or more on skill or practice. This will necessarily complicate the discussion in this chapter, since my argument pertains to testimony's (non-) propositionality, and this will be more or less interesting or persuasive depending on the degree to which you think propositional knowledge is involved in procedural knowledge. If you take the paradigm case of knowledge how to be something like "S knows how to prove the Pythagorean Theorem," then the argument that we can transmit knowledge-how through testimony will strike you as much less interesting—and controversial—than if you take the paradigm case to be closer to "S knows how to knit."

The reason I bring this up is not to take a stance in this debate. In fact, I mean to bracket the tension between anti-intellectualism and intellectualism altogether for the time being. For this chapter, I am not directly concerned with knowledge-how, but with *testimony-how*. I focus on what constitutes testimony-how and leave the question of whether a subject knows-how on the basis of another's testimony-how as an avenue for further research (but see Carter and Pritchard 2015; Hawley 2010; Peet forthcoming; Poston 2009; 2016). While the anti-intellectualist/intellectualist issue is relevant to the topic of the present chapter, the divide is epistemological and not metaphysical. The debate can be therefore decided independently of my argument for testimony-how and my account of its nature.

The purpose of the chapter is to spell out exactly how it is that one agent can learn how to φ from another; this is perfectly compatible with knowledge-how's being a species of knowledge-that. Moreover, since the intellectualist already holds that to know how to do something is to know some fact or set of facts about the world (Stanley 2011), he should agree that any utterance capable of instilling beliefs about the relevant facts should constitute testimony-how. In other words,

propositional statements about the relevant facts of how to do something (as the intellectualist understands those facts to be) just are, by the intellectualist's own lights, testimony-how. Nevertheless, the intellectualist still needs to explain how imperative utterances (namely, instructions) can convey facts. This chapter is meant to do just that.

However, to claim that one agent can learn how to φ from another does not commit one to the view that knowledge-how is a species of knowledge-that. Because my claim that an agent can come to learn how to φ from another does not turn on the claim that what is communicated is propositional knowledge, I do not commit to anti-intellectualism of either the strong or weak variety.

Because my argument is neutral with respect to the debate over the relationship between knowledge-that and knowledge-how, both the intellectualist and anti-intellectualist alike should be open to it. To say that there is such a thing as testimony-how neither precludes nor necessitates that testimony-how is a species of testimony-that, or vice versa.

2.2.2 A working definition of testimony

The following definition should not be confused with an *account* of testimony; indeed, developing an account is the project of Section 2.7 of this chapter. The purpose of the present section is to elucidate the general phenomenon at issue when epistemologists talk about testimony, in order to minimise confusion throughout this discussion. Part of this definition will be taken from Chapter 1 of this thesis, while other parts will be taken from the literature. It is, therefore, probably worth saying a few things that I already take for granted and are supported by the literature.

Crudely put, to learn from testimony is to learn from somebody's telling something to somebody else. One important feature I take testimony to have is that there need not be a strict 1:1 ratio between what is testified to and what is said; i.e. one can testify to more than what the literal content of the testimony conveys

(Lackey 2008), as when you ask me whether it is raining and I reply that there is an umbrella in the closet.

Second, and relatedly, testimony is constructive as opposed to transmissive, meaning that testimony is not a matter of implanting ideas from one person's mind into another's, but rather testifier and recipient alike perform cognitive labour in order to produce and interpret the content of the testimony. I argued in Chapter 1 that mindreading—in which testifiers and recipients must discern the other's motives, intentions, background knowledge, and how these are influenced by the context of the exchange—is a key way by which testifiers and recipients perform this labour. Moreover, this process contributes information that becomes part of the resultant beliefs.

Third, testimony can take many forms. Testimony can be spoken, signed, gestured, written, or otherwise offered; what is important is that there is some central act of communication that conveys communicable content. (I discuss this at length in Chapter 1.) This will mean, for the argument presented here, that if we understand instructions as being a kind of testimony-how (as I argue we should), then there will be several different ways of instantiating that testimony. Demonstrations, written procedures, and spoken instructions will all constitute testimony-how, on my view, provided they meet the other conditions I discuss later in this chapter.

Fourth, I deny that a testifier must know in order to testify, and, correspondingly, a recipient need not gain knowledge in order to have learned from testimony.[2] A testifier might not know something (because he lacks justification or belief, or holds a false belief) and yet testify to it. Testimony, then, can generate knowledge, either by instilling a true belief in a recipient who already holds some justification, or by providing justification for a true belief the recipient already

[2] My view here is informed by Goldberg (2005), Graham (2000), and Lackey (1999; 2006b; 2008).

holds.[3] I therefore deny that testimony is a matter of knowledge or belief transfer; rather, the way epistemologists ought to understand testimony is as one person's instilling a belief in another.

Testimony is epistemologically unique in that it is generally understood necessarily to take a propositional form. The same is not true, however, for at least one other source of knowledge: perception. When I perceive that the table before me is red or that my dog's fur is soft, I do not perceive those as propositions; rather, I have a qualitative experience and infer from the experience the relevant proposition, such as "the table is red" or "the fur is soft," and so on.[4] The question, then, is what we really mean when we talk about testimony: should we understand testimony as being, broadly, anytime we learn from others? Or should we understand it as learning only from others' propositional utterances?

Let us entertain the former: it would imply that if (unbeknownst to you) I watch you perform an action, and then I successfully copy you, and in the process come to know how to perform the same action, it does seem as though I have learned from you. Or, at least, my knowing how to perform the action is counterfactually dependent on my witnessing your performing the action. But intuitively, this should not count as testimony: you were completely unaware that you were being observed, and your performing the action was not meant to communicate anything. So, treating the former as though it is testimony seems to miss the point about what testimony *is*.

Let us now entertain the latter: it would imply that you can know from others only when they put their knowledge into propositional form. But this seems too restrictive. It would exclude a case where I ask my friend where he put the book he borrowed from me, and he points to the shelf where the book sits. His utterance (if

[3] This is a view defended by Lackey (2005; 2007) and Graham (2006), in response to Senor (2007).
[4] See Sant'Anna (2018b) for a critical discussion of the arguments for understanding mental content as propositional.

his gesture is even properly understood as an utterance) has not taken a propositional form, or at least it is no more propositional than seeing a green apple and coming to entertain the proposition "this apple is green." Nevertheless, I come to know from him because he has expressed to me some information in an attempt to instil a certain belief in me, i.e. that the book is on the shelf. This is not merely perceptual, since I did not see the book on the shelf on my own. My belief that the book is on the shelf is counterfactually dependent on my friend's pointing at the shelf; i.e. it depends on his communicative action meant to instil a certain belief in me.

Above, I asked whether we should understand testimony as being, broadly, anytime we learn from others, or whether we should understand it as learning only from others' propositional utterances. From the examples above, it looks as though the former is too permissive, but the latter too restrictive. Testimony, then, is not simply learning from someone, but neither is it a matter of learning strictly from another's propositional utterances. The truth is probably somewhere in the middle. I return to both points throughout the chapter, but for now, I will close this section with the following remark.

Since the purpose of this chapter is to discuss a non-propositional form of testimony, and the existing literature overwhelmingly describes testimony as being propositional, it is necessary to propose a working definition of testimony. For obvious reasons, I want to resist the popular formulation "*S* testifies that *p*." I develop a more precise account of testimony-how in Section 2.7 of this chapter, but for the time being, I will say roughly that testimony is what happens when one agent instils beliefs[5] in another, as a result of that agent's intentional act of communication. I leave open the question of whether what is communicated is information that *p* or some non-propositional procedural information.

[5] The beliefs formed need not be propositional, even if the corresponding testimonial utterance is.

2.3. Commitment (i): All instructions are testimony

Let us revisit the inconsistent triad I advanced above. The first solution is to deny (i): that all instructions are testimony. In this section, I argue that instructions are a kind of testimony. It may seem that of the three triadic commitments, this one is the weakest. Why, then, should we not reject it out of hand? After all, doing so would resolve the inconsistent triad and leave us with two other *prima facie* entirely plausible and broadly accepted claims. But let me explain why (i) is not as far-fetched as it seems.

Recall from the working definition of testimony I set out in Section 2.2.2 that at its heart, testimony is a matter of one agent's instilling beliefs in another (or attempting to do so). So, instructions count as testimony if instructing is a way for one agent to instil beliefs in another. To reject triadic commitment (i), you would need to claim that (at least) some instructions are not testimony. The obvious way to do this is by finding a counterexample: a case where an instruction is delivered but it fails to qualify as testimony. To this end, let us consider different kinds of instruction.

Instructions take many forms. Hawley (2010) lists "a number of ways in which B can learn from A how to φ:[6]

1. A describes to B how to φ
2. A gives B imperative instructions how to φ ('do this, do that')
3. A describes to B how A φs (or something like φ)
4. B overhears A talking to someone else [C] about how to φ (or about how A φs)
5. A intentionally shows B how to φ, and B imitates A
6. B observes A φ -ing and imitates A
7. B observes A trying and failing to φ, and thereby works out how to φ (maybe A intends this, maybe not; maybe A thinks she knows how to φ, maybe not)

[6] Hawley uses the term "how to X" rather than "how to φ." For the sake of consistency, I have opted to substitute φ for X throughout this chapter.

8. Intentionally or not, A forces or encourages B to come to know how to φ (to use trial and error, to practice, to pay for lessons?)" (Hawley 2010, 400, list numbers added for clarity and ease of discussion)

I think it is clear in (1), (2), and (5) in Hawley's list that A is instructing B in how to φ. The instructional status of (3), I think, depends on whether A describes to B how A does φ with the goal of having B learn how to φ; it seems to me that iff A has this goal, and is counting on B to update indexicals in the appropriate way (i.e. to update "I [A]" to mean "I [B]"), then A is instructing B in how to φ. Describing (4), (6), (7), and (8) as instructions is more strained: in (4), A may be instructing C, but although B learns from A how to φ, A is not instructing B. In (6) and (7), A is not so much instructing B as A is simply φ -ing, and B learns from A how to φ through non-instructional means. Finally, for (8), that A is a participant in the exchange seems almost inconsequential; to my mind, it is more appropriate to say that A is operative in B's learning how to φ, than it is to say that B is learning how to φ from A.

Importantly, to say that B has learned how to φ from A is *not* to say that A has instructed B. But there are some cases in which B has learned how to φ from A, and it has been through A's instruction. Since the purpose of this chapter is to explore the epistemology of instructions, I will concentrate my attention on Hawley's (1), (2), the relevant kinds of (3), and (5) throughout this chapter.

To that end, let us consider candidate counterexamples to the first triadic commitment. A counterexample could take any of the following forms of instruction:

1. A describes to B how to φ
2. A gives B imperative instructions how to φ ('do this, do that')
3. A describes to B how A does φ (or something like φ)
5. A intentionally shows B how to φ, and B imitates A

What we need is a case where A performs one of the given actions without actually instructing B. The obvious target is a case in which A performs one of the given actions, but B fails to learn from A; this, however, is not a genuine counterexample. Recall from Chapter 1 that Lackey (2008) distinguishes between s-testimony and h-

testimony, where s-testimony is the speaker's intentional expression of communicable content conveying the information that *p*, and h-testimony is the hearer's reasonably taking an intentional expression of communicable content to convey the information that *p*. Because Lackey's view is disjunctive, it is possible for a testifier to attempt to instill a belief in a recipient and fail to do so, owing to misunderstanding, noncomprehension, obstinance, or some other obstacle. The attempt, however, is still testimonial. Similarly, if A performs an action meant to instruct B in how to φ, B's failure to learn how to φ is not reason to say that A's action was not instruction. Rather, we have something akin to Lackey's s-testimony without h-testimony.

Now that the issue of counterexamples has been put to rest, of the instructional forms in Hawley's list, (2) is the kind of instruction at issue as regards the inconsistent triad I have presented above. So, as I discuss the inconsistent triad in the coming few sections, the reader should bear cases of type (2) in mind; these are non-propositional imperative instructions. I return to cases of types (1), (3) and (5) in Section 2.7 of the chapter, but for now, what is important about it is that there is at least one case where B can learn how to φ from A through A's use of imperatives. It is, as Hawley emphasises (2010, 398), an empirical question as to how many of our instructions are offered in each of the forms she enumerates. Similarly, I wish to emphasise that the question here is *not* about which instructional method is the most successful, expedient, or psychologically agreeable for an instructee; what matters is that they are on an epistemic par with respect to their ability to convey information to a recipient.

Since nothing epistemologically significant seems to rest on an instructor's choice of instructional form, the forms of instruction I have followed Hawley (2010) in identifying are on an epistemic par: no matter the grammatical form of A's instruction, B can still come to know how to φ. To reject certain instructions on the grounds that they are not propositional is an *ad hoc* solution, apparently solely for

the purpose of maintaining a commitment to testimony's necessarily being propositional. To my mind, this commitment is unjustified. As I argued in Chapter 1 of this thesis, memory and testimony are analogous, and moreover, there is a longstanding picture of memory and perception as analogous.[7] Both the philosophy of perception and the philosophy of memory include debates over the contents of perception and memory; as far as I can tell, there is no corresponding debate in testimony. That testimonial utterances ordinarily take propositional form seems to have served as evidence enough that testimony just *is* propositional; I return to this point and argue explicitly against it in Section 2.5 of this chapter.

Ultimately, it seems that rejecting the claim that all instructions are testimony (in virtue of their non-propositionality) is not the way out of the triad. To reject triadic commitment (i), you would need to reject that instructions of Hawley's form (1), in which A describes to B how to φ, count as testimony. But clearly, instructions of form (1) *are* testimony: A describes to B how to φ, and B comes to know how to φ. Since nothing epistemologically significant hangs on whether A uses this formulation or another, such as form (2), where A provides imperative instructions to B as to how to φ, we cannot reject (2) as a form of testimony without also rejecting (1). And again, the only reason you would reject those non-propositional forms is if you had already accepted that testimony is necessarily propositional. As I will show in Section 2.5, however, this notion is unjustified, and rejecting (i) will cause you to run headlong into the larger question in the epistemology of testimony as to whether sources of propositional knowledge need themselves take propositional form.

Bracketing this latter question, which I answer negatively in Section 2.5, rejecting (i) is not a satisfactory solution to the triad. Nevertheless, suppose you are unwilling to reject it. You have an alternative: you might claim that all instructions are propositional. This amounts to the denial of triadic commitment (ii), which holds

[7] See Green (2006) on the epistemic parity of testimony, memory, and perception, and Sant'Anna (2018a) for an extended treatment of issues in the philosophy of memory and perception.

that some instructions are non-propositional. Let us consider this view next and investigate what denying it would mean.

2.4. Commitment (ii): Some instructions are non-propositional

The second *prima facie* plausible triadic commitment is that some instructions are non-propositional. The present section will be brief, for I take it as obvious that some instructions are non-propositional. Consider instructions given in the imperative mood, such as "Loop the yarn counter-clockwise around the right needle," "Press Ctrl+Alt+Delete twice to restart your computer," or "Pull the oxygen mask toward you, place it over your mouth and nose, and secure the strap over the back of your head." Sentences of this sort are endlessly producible, but what they all have in common is that they are not truth-apt; i.e. they do not have truth conditions. Instructions in the imperative mood cannot be true or false; "to ask whether [imperatives] are true or false seems without any sense" (Jörgensen 1937, 289).

We regularly instruct in the imperative mood, so it is at least *prima facie* plausible that some instructions are non-propositional. "Keep your hands at 10 and 2" is an instruction as to where to place one's hands on a car's steering wheel, but is non-propositional and therefore not truth-apt. So, it seems that at least some instructions are non-propositional.

One might, however, find this basic claim to be unpalatable, and so reject it. This brings us to the second possible solution to the inconsistent triad. Denying that some instructions are non-propositional amounts to some version of imperative cognitivism (henceforth, simply "cognitivism"). The core tenet of cognitivism is that all imperatives (and therefore instructions delivered in the imperative mood) express propositions. In this section, I consider different kinds of cognitivism and their strengths, but ultimately deny that rejecting (ii) is the best solution to the triad.

Since cognitivists claim that all imperatives express propositions, they hold that all imperatives have truth conditions. On this view, although instructions are

delivered in the imperative mood, they in fact contain disguised propositions; this implies that instructions straightforwardly count as testimony. This solution will require commitment to a translation schema. Imperative translation schemata are meant to express what is *contained in the meaning* of imperatives; their role is to capture exactly, completely, and only what is meant by a particular imperative utterance. In this chapter, I consider five prominent translation schemata: Reports Theory, Desires Theory, Deontic Theory, Predictions Theory, or Elliptic Theory (Clark-Younger 2014).[8] Clark-Younger (2014) explains how each translation schema treats imperatives (usually of the command type); below, I adapt each one for instructions.

Under Reports Theory (Jackson and Pettit 1998; Lewis 1970), imperatives are identical to reports of those imperatives. So, for A to instruct B to "Pedal!" is identical to A's saying "I instruct that you pedal!" On this view, "Pedal!" is true iff A so instructs B. Desires Theory (Hare 1952), in contrast, holds that imperatives express a speaker's desire that a certain state of affairs be realised. For A to instruct B to "Pedal!" is identical to A's saying to B "I desire that you pedal!" "Pedal!" is true iff A really does desire that B pedal. Under Deontic Theory (Hamblin 1987), imperatives express "you should" statements. "Pedal!" is identical to "You should pedal!" so "Pedal" is true iff it is the case that B ought to pedal. Predictions Theory (Gibbons 1960) holds that imperatives express predictions about the future. "Pedal!" is identical to "You will pedal!" so "Pedal!" is true iff B will pedal. Finally, according to Elliptic Theory (Anderson and Moore 1957; Bohnert 1945), imperatives are shorthand for a declarative sentence that is the disjunction of two states of affairs. As Clark-Younger (2014, 67) puts it, "[Elliptic] Theory holds that commands are ellipses of disjunctive sentences of the form 'you do x or else...' The translation of 'get off my

[8] These schemata apply to all imperatives, including commands, advice, instructions, and so on. I am only interested in instructions, which are grammatically identical to other kinds of imperatives, but have different pragmatics (Kaufmann 2012). Thus, even though a certain translation schema may suit some kinds of imperatives, I will go on to argue that none is suitable for instructions.

lawn!' would be, perhaps, 'either you get off my lawn or I'll call the police,' or more generally 'either you get off my lawn or y' where y stands for some unspecified unpleasant consequence. This will perhaps be a threat, but it might not be. This means that commands are true if either you obey them, or if the bad thing that was implicit in the command happens, or, presumably, both." So, "Pedal!" is identical to "Either you pedal or *y*," where *y* might be "the bike will not move" or "you will fall over," or something along those lines, and is true iff one of the disjuncts is true.

2.4.1 Reasons to be sceptical of imperative cognitivism *simpliciter*

Cognitivism is *prima facie* a very attractive solution: it smoothly renders non-propositional utterances into propositional statements with clear truth conditions. This is true regardless of the translation schema that we adopt. However, since I am about to argue that cognitivism will not provide the solution we need irrespective of the translation schema, I do not commit to a particular schema.

Cognitivism fails to solve the problem on three counts. First, there are general reasons to be sceptical of cognitivism's ability to explain even paradigmatic imperatives such as commands. Second, even if those concerns were assuaged, cognitivism struggles to explain instructions, largely because it struggles to explain soft imperatives of any kind. Third, even if it did not struggle to explain soft imperatives, cognitivism would still fail to provide an adequate account of how an instructee comes to gain beliefs about how to do things on the basis of a testifier's word. Cognitivism, then, is triply damned.

To explain why this is the case, let us begin with the general criticisms against cognitivism. The short version of the criticism is that cognitivism results in what Parsons (2012) calls "unwanted validities" and "unwanted consistencies," i.e. if imperatives are understood as having truth conditions and are incorporated into arguments of standard form (such as *modus ponens*), they sometimes come out as valid or consistent when they should not.

To see what Parsons (2012) means, let us consider a case where A instructs B. Suppose A is teaching B how to float on her back in a swimming pool and says "Push your bellybutton upwards." Let us consider, like Parsons, Reports Theory, on which this instruction would be equivalent to "I instruct that you push your bellybutton upwards." But consider the following pair of instructions:

A1. Push your bellybutton upwards.
A2. Do not push your bellybutton upwards.

A1 and A2 are clearly inconsistent: an instructee who received these instructions would, quite rightly, not know what to do. But the translations that result from Reports Theory do not bear this inconsistency. They would translate as follows:

B1. I instruct that you push your bellybutton upwards.
B2. I instruct that you do not push your bellybutton upwards.

But B1 is not inconsistent with B2 in the way that A1 is inconsistent with A2. One can coherently report that one has given inconsistent instructions, which suggests that imperatives are not in fact identical to reports of those imperatives. The problem is not unique to Reports Theory, either: see Clark-Younger (2014) for a detailed account of how the problem of unwanted consistencies arises even if we endorse a different translation schema.

Let us also consider the problem of unwanted validities, again using A1 as our example.

A1. Push your bellybutton upwards.
Therefore, someone instructs something.

On Reports Theory, this would translate as:

C1. I instruct that you push your bellybutton upwards.
Therefore, someone instructs something.

The latter of these arguments is valid: it cannot possibly be the case that "I instruct that…" is true without forcing the conclusion to follow. And if cognitivism is true,

then the former is also valid. But the conclusion seems not to be about the same thing as premise C1: the first is about pushing one's bellybutton upwards, while the second is about instruction. They lack the sort of relevance required for validity (Parsons 2012). So, this suggests that cognitivism fails.

Clark-Younger (2014) introduces two other problems for cognitivism, but for the sake of brevity, I pass over a good deal of her expansion on Parsons' (2012) view. I do, however, encourage the interested reader to consult Clark-Younger (2014), especially Chapters 5 and 6. The upshot is that cognitivism, regardless of the translation schema we adopt, will suffer from one or both of the problems of unwanted validities or unwanted consistencies; this already serves as a persuasive reason to reject cognitivism.

2.4.2 Reasons to be sceptical about specific translation schemata

But suppose that you are not convinced by what I have presented above: suppose, for whatever reason, you still think cognitivism is the way to go. You might think this is because the arguments I have considered so far are meant to apply to all imperatives, but especially orders and commands. But instructions are not interchangeable with imperatives; rather, they are a *species* of imperative, and exhibit certain features that set them apart from other kinds of imperatives. For one thing, instructions are a species of *soft imperative*, and along with requests and advice, bear an extra layer of difficulties when it comes to understanding what their propositional content is (Clark-Younger 2014). Extant translation schemata, though they are meant to explain imperatives and should in principle also explain instructions, fail to account for instructions.

Reports Theory

To see how the translation schemata fail to provide adequate accounts of instructions, let us begin with Reports Theory, under which instructions are identical to reports of those instructions. If A says to B, "Pedal!" then "I instruct that you

pedal!" is true iff A really is *instructing* B on how to ride a bike (or how to move forward, or how to climb to the top of the hill, and so on). But A's utterance of "Pedal!" could be a number of different types of speech act: she could be *commanding* that B pedal (if A is, for example, B's cycling coach), she could be *requesting* that B pedal (if A and B are riding a tandem bicycle and B is doing all the work), or she might be *advising* B to pedal (if B is coasting but would do better to continue pedalling). For a successful translation, Reports Theory requires that we know what type of communication A intends in order to determine the truth conditions of A's utterance.

This distinction might seem unimportant; after all, even in paradigmatic testimony, we can happen to gain true beliefs from a testifier even if we misunderstand the testifier's meaning. So it might not seem to matter whether B thinks A is instructing when she is really commanding, and so on. But if this is the case, then instructees are routinely gaining the wrong kinds of beliefs on the basis of an instructor's instructions. (I address this worry in more detail in Section 2.4.3.) For now, suffice it to say that a difficulty translating the correct speech act poses a significant enough challenge for Reports Theory that we should reject it as a candidate.

Desires Theory

Let us next consider Desires Theory, under which instructions express an instructor's desire. It is, however, not contained in the *meaning* of "Pedal!" that the instructor desire that the instructee pedal, so the translation is at best logically suspect. Furthermore, we might instruct people to do things that we do not actually desire that they do. For example, consider the wildly implausible (if dramatically compelling) scenario on the ABC drama *Private Practice*, in which pregnant medical doctor Violet Turner is assaulted by her former patient, Katie, who is convinced that the baby is rightly hers. When Katie (who has no medical training) begins to read a textbook to learn how to perform a C-section to retrieve the baby, Violet, realising

Katie fully intends to go through with it, insists that Katie follow her instructions so as to save the baby, even though Violet will likely not survive. In this scenario, Violet instructs Katie on how to perform a C-section even though Violet very clearly does not desire that Katie do so.

The reader may, however, object to this example on the grounds that Violet *does* desire that Katie follow her instructions to perform the C-section: she desires it as a means to save her baby when it becomes clear to her that Katie is going to operate on Violet one way or another. So, you might think that Violet has an adaptive preference, and in that sense, when she instructs Katie to "Make the first incision here," "I desire that you make the first incision here" is indeed the appropriate translation. I myself am not in this camp—I think it is fairly clear that under normal circumstances, Violet would not desire that Katie perform a C-section on her—but if the reader is convinced by the objection, let us consider an alternative. Suppose a YouTube video instructs viewers in how to change a flat bicycle tire. It is unlikely that the instructor has any desires regarding the viewer's changing the tire, and yet that does not impede the instructor's ability to instruct: what is more important is whether the instructor is giving good instructions, i.e. ones that, if followed, will lead the viewer to change a tire successfully. The first example is extreme, and the second is mundane, but the point stands: Desires Theory fails as a translation schema for instructions for the simple reason that instructions are not always given in accordance with an instructor's desire.

Deontic Theory

Let us now turn to Deontic Theory, under which imperatives can be understood as expressing "you should" statements; i.e. "Pedal!" translates as "You should pedal!" Clark-Younger argues that Deontic Theory fails, since "the 'shoulds' must be thought of as all-things considered shoulds" (2014, 82), but it is far from clear that if instructions generate obligations, they are all-things-considered obligations. Let us explore this idea in more detail. Deontic Theory is consistent with Lewis' (1969)

claim that imperatives generate an obligation, but that claim does not seem to translate well to instructions. Though commands or orders seem to generate obligations, instructions lack the same urgency and obligatory force. Furthermore, it is not contained in the meaning of the instruction "Do X" that an instructee is now *obligated* to do X. Instructions, unlike their command counterparts, are obligation-neutral. I can follow them or I can not, although my successful φ-ing may depend on my following the instruction to do X (and if there is an obligation that I φ, I may be obligated to do X, but the obligation does not come from the fact that I have been so instructed).

But Deontic Theory is concerned with permissions and obligations alike, so we might think that it will suit us to understand instructions as expressing permissions rather than obligations. Indeed, if I ask you how to φ, you might well reply with something like "You can do it this way," which quite clearly expresses a permission (and, moreover, is truth-apt). However, while it may be contained in the meaning of "Do X" that an instructee is *permitted* to do X, mere permission does not adequately capture the implicit assurance that doing X is a way to φ. Surely there are many things that I *can* do that are compatible with my φ-ing without being ways to φ: if I want to knit a lace shawl, I am permitted to drink a gallon of tea, to adjust the thermostat, to cook dinner, to feed my dog, to assemble a chest of drawers, to read philosophy, and so on, but none of these actions are ways for me to knit a lace shawl. Understanding instructions as expressing permissions, then, provides at best an incomplete solution: we need a way to capture the implicit guarantee that doing X is a way to φ.

On some accounts of paradigmatic testimony (i.e. testimony-that), testifying involves an implicit invitation to trust (e.g., Hinchman 2005). When testifiers testify, they extend an invitation to a recipient to trust that what the testifier says is true. Of course, as we are well on the way to establishing by now, instructions in the imperative mood are not truth-apt, so the notion of an invitation to trust that what

the instructor says is *true* seems slightly odd. Nevertheless, it is not too difficult to imagine an analogue to sincerity. What it would look like in the instruction case? It would probably take a form like "Doing X is a way to φ [and my saying so is a guarantee that this is true]."

This looks awfully close to a solution, but will not get us where we need to be. The role of a translation schema is to capture what is contained in the *meaning* of the instruction, and it is far from clear that the guarantee is contained therein. Rather, the guarantee acts more as an add-on to the meaning of the sentence, where the sentence itself means something like "You are permitted to do X" and the fact of the instructor's saying so acts as the guarantee. For Deontic Theory (and indeed, any of the translation schemata) to be satisfactory, it must capture exactly and only what is contained in the meaning of the instruction. If the point of the translation schema is to translate a truth-inapt sentence to a truth-apt one, then no other changes should be acceptable; to impose the guarantee clause to make the permissions variant of Deontic Theory palatable is not the right way to go. Moreover, when we couple this failure with the challenges that the problems of unwanted consistencies and unwanted validities pose to Deontic Theory, we must conclude that the solution does not lie therein.

Predictions Theory

So much for Deontic Theory; perhaps the solution lies in Predictions Theory, which holds that instructions express an instructor's prediction that the instructee will follow her instructions. On this view, "Pedal!" is identical to "I predict that you will pedal!" But it is not the case that instructors always predict that their instructees will follow the instructions. I might instruct an undergraduate philosophy class on how to complete a truth table, but that does not mean that I predict they will all successfully learn how to complete a truth table immediately after I teach them; in fact, I may predict that most of them will *not* immediately learn how to complete a

truth table purely on the basis of my instruction. (In addition to instruction, they will need practice, for example.)

One way out of this, you might think, is to deviate from instructions for a moment and return to commands, another species of imperative, and say that commands issued contrary to a speaker's predictions simply do not count as genuine commands; i.e. if a speaker commands something while predicting that the command will not be followed, then the speaker has not genuinely commanded anything. You might think that instead the speaker has merely expressed a wish, for instance. The problem is that the same solution does not seem to hold for instructions (Clark-Younger 2014). There seems to be no important reason that a speaker's predictions have anything to do with the ontology of instructions and whether a given utterance counts as an instruction. At best, if I instruct you to do something, I might be predicting (and hoping) that you come to *know how* to do that thing, but I am not necessarily predicting that you will *actually* do it.

Part of the reason that Predictions Theory fails is that the translations do not distinguish between types of imperative: "Pedal!" may be an instruction, a command, a request, a wish, or so on, but no matter which one it is, it will be translated as "I predict that you will pedal." However, the differences among types of imperative are significant and any adequate translation schema should be able to account for those differences. Treating them as though they were all the same just means that our translation fails to capture something contained in the meaning of any given imperative. Furthermore, this last point holds true for all other translation schemata that we have considered so far except for Reports Theory, which itself runs into the problem of being able to identify what speech act is being performed in the utterance of any given imperative sentence.

Elliptic Theory

Let us finally consider Elliptic Theory. Elliptic Theory seeks to capture the "motivating force" of commands by understanding imperatives as disjunctions, where one disjunct is a prediction of the addressee's action and the other is a prediction of the (usually negative) consequence of the addressee's failure to fulfil the command. So, "Pedal!" becomes "Either you will pedal or you will fall over!" (for example).

The main problem here is that Elliptic Theory introduces much more information than is in the command. Nothing about "Pedal!" implies that "Either you will pedal or you will fall over!" The information comes from the context but is not contained in the meaning of the sentence itself (Clark-Younger 2014). So, in addition to introducing extra information, it also runs the risk of introducing extra *false* information. You might not think this is a problem; after all, I have previously followed Lackey (2008) in arguing that there need not be a strict 1:1 ratio as to what is uttered and what is testified to. For instance, if you ask me if it is raining, and I reply that there is an umbrella in the closet, I have implicitly testified that it is indeed raining while also explicitly testifying that there is an umbrella in the closet.[9] Perhaps, you might think, the same can be true of imperatives, and rejecting Elliptic Theory on the grounds that it introduces information not contained in the meaning of the sentence is inconsistent with a view of testimony that allows that recipients can understand more from a testifier than what she explicitly says. You may furthermore claim that the ability for Elliptic Theory to introduce *false* information is not a problem because testifying is not factive.

This objection is misguided because of the role that logical translation plays and the role of the recipient's faculty of comprehension. There is a difference between the translation of an utterance and what a recipient might understand from

[9] This example is borrowed from Lackey (2008).

that same utterance. Translation must be exact; it's not about inference or what follows from a statement. When a recipient interprets a piece of testimony, she might interpret in such a way that she introduces new information based on her background knowledge, the context of the exchange, or any of the other factors I addressed in Chapter 1. But this is not what translation is: translation ought to deal in only that information that is (logically) contained in the utterance. It is perfectly fine for an addressee to introduce new information psychologically, but a mistake for a translation schema to do the same.

Even if you are unsatisfied by my response to this objection, it is worth noting the clear relationship between Elliptic Theory and Predictions Theory: Elliptic Theory is simply a matter of making a disjunctive prediction. As such, it will suffer from the same problems as Predictions Theory. Most significantly, the introduction of other information into the meaning of the instruction is a major worry, and since it is a ground to reject Predictions Theory, so too should it be a ground for rejecting Elliptic Theory.

2.4.3 Reasons to be sceptical of cognitivism about instructions

As Clark-Younger (2014) argues, the problems of soft imperatives, unwanted consistencies, and unwanted validities should lead us to conclude that cognitivism does not provide an adequate account of imperatives. We should, therefore, be sceptical about cognitivism in general.

But suppose that the general reasons to be sceptical of cognitivism are not convincing in the case of learning from instructions. Suppose that you think, for whatever reason, that the objections to cognitivism fail and that cognitivism serves as an adequate explanation of instructions. You would then confront the consequence that one of two (equally implausible) things is happening if instructions are properly understood as declaratives in disguise. One possibility is that each imperative seems to constitute testimony about the wrong kind of thing: "Pedal!"

does not translate in such a way that it imparts propositions about how to pedal, but instead about the mental states of the instructor or her assertions about the world. The implication is that instructees are gaining the wrong kinds of beliefs (e.g., "Chloe wants me to pedal like this" rather than "This is how to pedal"). So, "Pedal!" conveys information that "Chloe wants me to pedal like this" when we think it is conveying information about how to pedal.

The other possibility is that instructees are routinely failing to gain the knowledge expressed: the instruction is stated as an imperative, translated to a declarative, and then taken up as an imperative once more. This process is implausibly cumbersome; furthermore, it would indicate the routine failure of comprehensive capacities, since instructees would be required to gain knowledge of how to pedal ("Pedal like this") even though the translation of the imperative is something like "I want you to pedal like this." The simpler and more cognitively economical explanation is that instructees simply take up instructions in their given imperative form. If the solution to the puzzle lies in adopting cognitivism about instructions, it does not reside in these translation schemata.

To sum up, the incorrectness of cognitivism about instructions is overdetermined. First, we have general reasons to be sceptical about cognitivism; second, we have specific reasons to be sceptical about cognitivism's ability to account for instructions *qua* species of imperative; and third, we have even more specific reasons to be sceptical about cognitivism's fit with how instructees form beliefs on the basis of an instructor's word.

The whole discussion about cognitivism notwithstanding, however, another reason that denying (ii) will not solve the inconsistent triad is that instructions simply do not always come in the imperative mood. We can instil knowledge-how in other people through other means; for example, sometimes we use the indicative mood, but "if B asks A how to balance on one foot without wobbling, nothing rests on whether A responds 'you should put your finger on your nose' [indicative] or just

'put your finger on your nose' [imperative]. Either way, B now knows what (s)he did not know before, namely, how to balance on one foot without wobbling" (Hawley 2010, 400–401). In other words, exactly the same procedural knowledge can be imparted through an imperative or an indicative. That's not a problem for cognitivism, which is only concerned about the relationship between imperatives and indicatives, but still poses a problem for us when it comes to figuring out exactly how it is that indicatives contain procedural information.

However, even if we could account for the apparent insignificance of whether instructions are delivered as indicatives or imperatives, there is yet another layer of complication: not all ways of imparting knowledge-how are verbal, let alone categorisable as indicative or imperative. One notable example is when we tell people how to do things by showing, demonstrating, or ostending, where—to borrow Hawley's example—A, when B asks how to balance on one foot with one wobbling, might answer simply by putting her finger on her nose. This kind of instruction is nonverbal and yet is clearly capable of instilling knowledge-how: B now knows how to balance on one foot without wobbling because A showed her how to do so.

If we claim that imperatives really are propositional by accepting cognitivism, we run into at least one of three problems. Either we encounter the problems of unwanted validities and unwanted consistencies (Clark-Younger 2014; Parsons 2012), we have to account for the apparent failure of cognitivism to explain instructions *qua* species of imperative, or we have to accept that instructees are routinely failing to gain the correct beliefs about what is being expressed. None is desirable. Furthermore, even if we accepted cognitivism, there are still epistemologically fuzzy cases that would fill the shoes of ordinary imperative instructions, and pose analogous problems. For example, cognitivism would be unable to account for things like demonstration, which conveys information that can be put in propositional form, but is not itself propositional. This leaves us exactly

where we started with a kind of instruction that is non-propositional. So, denying (ii), that instructions are non-propositional, will simply not get us to where we need to be, and will not present a viable solution to the triad.

In the next section, I consider triadic commitment (iii): that all testimony is propositional.

2.5. Commitment (iii): All testimony is propositional

Like Sections 2.3 and 2.4, this section is devoted to a triadic commitment. In this case, the focus is on (iii), the claim that all testimony is propositional, which accords with the dominant voice in the literature. Received wisdom has it that testimony is, roughly, asserting a proposition as true (Adler 2017). The assumption that testimony is propositional is understandable but misguided; nevertheless, it has been influential in the evolution of the epistemology of testimony. Let us consider the following canonical accounts of testimony, all of which focus on propositional forms of testimony.[10] I set aside the question of whether each account is independently plausible, since the purpose of the present section is not to evaluate them but instead to show that the notion that testimony is propositional is common to all. It is worth noting, however, that there are compelling arguments for and against each account.[11]

Fricker (1995), Audi (1997), and Sosa (1991) offer accounts of testimony that are attractive in their simplicity: namely, they claim that a subject S testifies that proposition p iff S's statement that p is an expression of S's thought that p (Lackey 2008). Fricker puts it as "tellings generally" (1995, 397) but proceeds to explicate her argument in terms of telling that *p*. Audi describes testimony as "people's telling us

[10] The arguments I have rehearsed here do not exhaust the views on testimony. They are, nevertheless, canonical in the literature and are characteristic of the types of views that pervade the field. While it is possible that there is a non-propositional account of testimony of which I am unaware, the accounts I have presented here nevertheless illustrate the ubiquity of the view that testimony is propositional.

[11] For a helpful overview of the debate, see Lackey (2008), especially Chapter 1, Gelfert (2014), and Lackey & Sosa (2006).

things" (1997, 406), but he, too, proceeds to rely on the notion of propositions as the form of testimony. Certainly, even if he would be content to expand the notion such that testimony (or testimonial utterances) need not take propositional form, the resultant account of testimonial justification that he develops frames everything in terms of believing that *p*, so there is clearly a propositional sense of testimony at work in his definition. Sosa understands testimony as "a statement of someone's thoughts or beliefs" (1991, 219). Lackey groups these three views together to and says that what they commonly express is the view that "S testifies that *p* if and only if S's statement that *p* is an expression of S's thought that *p*" (2008, 20).

Unlike the accounts of Fricker, Audi, and Sosa, Coady's landmark account has an epistemic condition built into it. He argues that S testifies by making some statement that *p* iff:

i. "S's stating that p is evidence that p and is offered as evidence that p.
ii. S has the relevant competence, authority, or credentials to state truly that p.
iii. S's statement that p is relevant to some disputed or unresolved question (which may or may not be whether p) and is directed to those who are in need of evidence on the matter." (Coady 1992, 42)

In a similar vein, Graham posits an account of testimony in which S testifies by making some statement that p iff:

i. "S's stating that p is offered as evidence that p,
ii. S intends that his audience believe that he has the relevant competence, authority, or credentials to state truly that p,
iii. And S's statement that p is believed by S to be relevant to some question that he believes is disputed or unresolved (which may or may not be whether p) and is directed at those whom he believes to be in need of evidence on the matter." (Graham 1997, 227)

Here, too, I wish to emphasise the role of propositions in both Coady's and Graham's accounts: that the act of communication that matters is "S's stating that p." This kind of account seems to leave little room for non-propositional content to qualify as testimony. Moreover, that they have a built-in epistemic condition

challenges their viability as even an account of testimony-that, let alone testimony-how. (I return to this point in Section 2.6.)

By contrast, Lackey's account of testimony is unique in its dual attention to the respective and complementary roles of speaker and hearer. On her view, S testifies that *p* by making an act of communication *a* iff (in part) in virtue of *a*'s communicable content:

i. "S reasonably intends to convey the information that *p* or
ii. *a* is reasonably taken as conveying the information that *p*." (Lackey 2008, 35–36)

The act of communication in question, *a*, amounts to the "intentional expression of communicable content" (Lackey 2008, 28), and can convey that *p* in one of three ways. First, *a* may be the utterance of a declarative sentence such that it expresses the proposition that p. Second, *p* may be an obvious (uncancelled) pragmatic implication of *a*. And third, it may express the proposition that *q*, and may convey that *p* so long as it is obvious (either to everyone in the exchange or to a normal competent speaker) that *q* entails *p* (Lackey 2008, 29). What is interesting about this account is that Lackey does not require that the act of communication *take the form* that *p*; it need only *convey* that *p*. I take this to mean that certain non-propositional acts of communication (such as a gesture) that convey that *p* might still constitute testimony under the right conditions. Nevertheless, whether it is expressed propositionally or non-propositionally, the content of the testimony seems for Lackey to be necessarily propositional.

In the end, it probably does not matter too much which of these accounts one endorses (though I return to the question in Section 2.6 of this chapter). What is central to each account is that testimony is a matter of transferring (or reproducing, or instilling) beliefs in the form of truth-apt propositions. And though they do not, in principle, preclude the possibility of non-propositional testimony, the notion that there is such a thing appears to be little more than an afterthought or idle question

kicked somewhere further down the theoretical road (see, for example, Chakrabarti (1994) and Jack (1994), who each suggest that instructions or commands might count as testimony, without engaging with this idea in depth).

I suspect that the centering of propositionality in the epistemology of testimony is a self-perpetuating phenomenon, in that testimony is assumed to be propositional and then accounts of testimony and related phenomena inherit this assumption, to the point where testimony almost seems to be necessarily propositional and it is hard even to conceptualise what non-propositional testimony would be. Burge's "Acceptance Principle," for example, is that "a person is *a priori* entitled to accept a proposition that is presented as true and that is intelligible to him, unless there are stronger reasons not to do so" (1993, 467). (Note Burge's explicit focus on propositions.) The fixation on propositions probably goes some way toward explaining the intuitive oddness of the phrase "non-propositional testimony." The assumption, however, is mistaken.

So far, we have found triadic commitments (i) and (ii) to be true. This brings us to the third, and, as I will argue, most viable solution to the inconsistent triad: the denial of (iii), that all testimony is propositional. Here, I describe why one might have come to endorse (iii), and the reason is simple: namely, many canonical accounts of testimony assume it. Although these accounts are suffused with the assumption that testimony is propositional, I think that the assumption is worth questioning, and indeed, rejecting.

Before we can build an account of testimony that is neutral with respect to whether testimony is propositional, we need to ask what reasons we have to do away with the assumption that testimony is propositional. After all, this may seem like a rather *ad hoc* solution, presented solely in the interest of letting instructions into the testimony club. But I think there are compelling independent reasons to abandon the assumption that testimony is exclusively propositional.

I said above that it was an idiosyncrasy of testimony that it is taken to be necessarily propositional when other sources of propositional knowledge are not subject to the same restrictions. Even epistemologists who deny that perception is propositional still accept that perception can give us propositional knowledge. Why, then, is it so difficult to accept that testimony might also have non-propositional content? I also said above that there were such things as knowledge-how and knowledge-that, and that there was a way of instilling knowledge-that in another person. My opponent could then respond that testimony is just the way that agents instil propositional knowledge in other people. He may even grant that there is a way to instil knowledge-how in other people, but he may object to categorizing that way of instilling knowledge-how as *testimony*.

One possible response is to say that the only reason for someone to deny that testimony-how genuinely counts as testimony is if he has already bought the view that testimony is necessarily propositional. (I am about to deny that very view.) In other words, someone who simply denies that testimony-how is testimony is probably doing so because they already have an unreflective or implicit commitment to the propositionality of testimony. However, the unreflective commitment notwithstanding, there appears to be no principled reason to claim that testimony-how does not genuinely count as testimony. In the face of the evidence for triadic commitments (i) and (ii), and without a principled reason to defend (iii), it seems that the best way out of the inconsistent triad is to reject the claim that testimony is necessarily propositional.

2.6. Some instructions should count as testimony

If we were to suspend the notion that testimony is propositional and open the doors a little wider, I think we would build a much more robust account of testimony. Amending the conditions on testimony to include non-propositional utterances that convey information that the recipient did not previously know (or believe) results in

an account of testimony that more readily accommodates instances that intuitively constitute testimony without being too broad (i.e. it continues to exclude instances that ought not to count as testimony).

Recall that we started with this inconsistent triad:

i. All instructions are a kind of testimony.
ii. Some instructions are non-propositional.
iii. All testimony is propositional.

As I showed above, the best way out of the inconsistent triad is to reject (iii), that testimony is always propositional. Since we already know that we must accept (i),we can conclude that (ii) is true. In this section, I offer some further considerations in support of (ii), which will serve to reinforce my argument that there is such a thing as testimony-how.

Let us return to the standard propositional accounts of testimony that I addressed in the previous section. If we were to suspend the proposition condition each account posits (even if it does so only implicitly), we will see that instructions otherwise satisfy the requirements of each. This, I think, shows that instructions really ought to count as testimony.

Let us begin with the Fricker-Audi-Sosa view of testimony, which Lackey summarises as claiming that "S testifies that *p* if and only if S's statement that *p* is an expression of S's thought that *p*" (Lackey 2008, 20). As I have argued above, we ought to suspend the propositionality requirement. Let us therefore reformulate the view by replacing proposition *p* with some thought *t* about the world. In Lackey's formulation, this sounds odd, as it requires an expression of S's thought that *t*, but *t* is meant merely to signal a thought in either a non-propositional or propositional form. It becomes clearer when we consider the individual views, and not just the summary of each. The rest of the section is devoted to replacing uses of *p* in the standard accounts of testimony with *t*. Let us begin with Fricker (1995).

Recall that Fricker describes testimony as "tellings generally" (1995, 397). The language of *p* continues to play a large role in what follows of her argument. Slotting in *t* does not seem to pose any problem for Fricker's view: tellings can certainly be thoughts about the world. For Audi, on the other hand, testimony is "people's telling us things" (1997, 406) and results in "knowledge […] received by transmission" (1997, 410). Surely for him too, however, the things that people might tell us could be thoughts about the world. Finally, Sosa understands testimony as "a statement of someone's thoughts or beliefs" (1991, 219). Again, there is no conflict with replacing the propositions Sosa has in mind with thoughts about the world.

Why, then, should we not simply adopt one of these accounts, since they can so readily accommodate testimony-how? The answer is that there are many independent reasons to reject any of these accounts, so it is not an indication that reformulating them with *t* in place of *p* will get us very far as an account of testimony-how after all. For instance, the problem with any of the Fricker-Audi-Sosa views, as Lackey (2008) has pointed out, is that they are mind-bogglingly broad. Nearly every utterance gets in the door: idle wonderings, empty remarks about the weather, and talking to oneself count as testimony right alongside giving directions to strangers, testifying in court, and giving a lecture.

It is therefore simply unremarkable that instructions count as testimony, for these views set the testimonial bar too low to begin with. So, it is all well and good that for any of these accounts, *t* fits into the formulations in *p*'s place, but those won't serve as adequate accounts of testimony-how in virtue of the fact that they don't serve as adequate accounts of testimony at all. Nevertheless, for those who are unconvinced by Lackey's objections to the Fricker-Audi-Sosa view, we can end here, satisfied that such a view offers an acceptable account of testimony-how.

But for those convinced by the objections to the Fricker-Audi-Sosa view, and who are dissatisfied by the prospect of putting testimony-how in those terms, let us

next consider Coady's (1992) account. Recall that Coady argues that S testifies by making some statement that *p* iff:

i. "S's stating that p is evidence that p and is offered as evidence that p.
ii. S has the relevant competence, authority, or credentials to state truly that p.
iii. S's statement that p is relevant to some disputed or unresolved question (which may or may not be whether p) and is directed to those who are in need of evidence on the matter." (Coady 1992, 42)

Similarly, Graham argues that S testifies by making some statement that p iff:

i. "S's stating that p is offered as evidence that p,
ii. S intends that his audience believe that he has the relevant competence, authority, or credentials to state truly that p,
iii. And S's statement that p is believed by S to be relevant to some question that he believes is disputed or unresolved (which may or may not be whether p) and is directed at those whom he believes to be in need of evidence on the matter." (Graham 1997, 227)

What sets Coady's and Graham's views apart from others is that both impose an evidentiary requirement. In turn, this means that both Coady's and Graham's views alike are subject to the same criticism: that the imposition of this evidentiary requirement makes the account an epistemological one rather than an ontological one. By conflating testimony's epistemology with its ontology, we lose the ability to distinguish the question of what testimony *is* from the question of whether a testimonial recipient knows that p on the basis of S's testimony.

Furthermore, as an account of testimony-how, the Coady-Graham view will struggle to accommodate instructions in imperative form because of the evidentiary requirement. But this, you might object, is not evidence against the Coady-Graham view; instead, it is evidence that instructions simply should not count as testimony. I disagree, on the ground that there are many types of utterances that are (intuitively) testimony that will not count as testimony on the Coady-Graham view. Furthermore, the Coady-Graham view precludes false testimony from counting as testimony at all, because false testimony fails to meet the evidentiary requirement. This is a prime

example of a time when the conflation of the question of what testimony is with the question of what good vs. bad testimony is rears its ugly head. Because this view precludes such utterances, we have independent reason to reject it, so the fact that it struggles to accommodate instructions should not worry us: rather, it is a downstream consequence of a view that fails as an account of testimony simpliciter.

Finally, Lackey (2008, 35–36) argues that "S testifies that *p* by making an act of communication *a* iff (in part) in virtue of *a*'s communicable content:

i. S reasonably intends to convey the information that *p* or
ii. *a* is reasonably taken as conveying the information that *p*."

On this view, *a*, the act of communication in question, is the "intentional expression of communicable content" (Lackey 2008, 28) and conveys that *p* (Lackey 2008, 29). This is the best general account of testimony because it avoids the undesirable breadth of the Fricker-Audi-Sosa view. It also avoids the evidentiary requirement that makes the Coady-Graham view untenable, and thereby allows us to distinguish the question of what testimony is from the question of what differentiates good testimony from bad testimony. Finally, Lackey's view takes into account that both testifier and recipient are participants in testimonial exchanges, and in doing so enables us to understand testimony both as source of belief and as communication of information. In the next section, we will turn to Lackey's account as a source for a broader account of testimony that can accommodate testimony-how.

2.7. Instructions as testimony-how

The chapter so far has been dedicated to defending the claim that testimony is not necessarily propositional, and that instructions are the typical case of testimony-how. But I wish to explicate the notion of testimony-how further, and provide an account of testimony-how. I propose the following account of testimony-how: *S testifies how to φ if S performs an act of communication that conveys information about how to φ*. The influence of Lackey's (2008) view of testimony should be clear, especially

given what I said in Sections 2.5 and 2.6. The key difference is that where Lackey says that testifying depends on either S's reasonably intending to convey the information that *p* or *a*'s reasonably being taken to convey the information that *p*, I take the relevant thing to be S's conveying information about *how to φ*, or *a*'s reasonably being taken to convey information about how to φ. Put in the proposition-neutral terms I used in Section 6, S testifies *t*, where *t* is some thought about the world.

2.7.1 Toward an account of testimony-how

This invites us to revisit Lackey's s-testimony and h-testimony. Recall from Chapter 1 and Section 2.3 of this chapter that Lackey's view of testimony is disjunctive. For Lackey, testimony can be either an intentional act on the part of the testifier, or a source of belief for a recipient. In this case, since we are concerned with instructions, an instructor testifies if she reasonably intends to convey information about how to φ, the testifier will be an instructor, and the recipient an instructee. So, let us adapt Lackey's view, on which

> "S testifies that *p* by making an act of communication *a* iff (in part) in virtue of *a*'s communicable content:
>
> i. S reasonably intends to convey the information that *p* or
> ii. *a* is reasonably taken as conveying the information that *p*." (Lackey 2008, 35–36)

Let us adjust it for instructions thus:

> S testifies *t* by making an act of communication *a* iff (in part) in virtue of *a*'s communicable content:
>
> i. S reasonably intends to convey the information *t* or
> ii. *a* is reasonably taken as conveying the information *t*.

This formulation of testimony-how allows us to include failed instructions, i.e. instructions where the recipient has not gained beliefs about how to φ, as well as peculiar cases, which I address below.

Above, I discussed imperative instructions as the paradigm case of testimony-how. But the view of testimony I posited above allows us to account for cases of many kinds other than imperative instructions, such as Hawley's (1), (3), and (5), reproduced here:

1. A describes to B how to φ
3. A describes to B how A φs (or does something like φ-ing)
5. A intentionally shows B how to φ, and B imitates A.

Let us consider each of these cases in turn.

Case (1) should obviously count as testimony, since A is likely to offer her description in propositional form. But whether this testimony counts as testimony-how or testimony-that is somewhat less obvious, for it seems clear that A's testimony is *about* how to do something, but less clear that it counts as testimony-*how* in the relevant way. In either case, the expanded, proposition-neutral account of testimony can accommodate Case (1), and we can leave the question of whether it is better understood as testimony-how or testimony-that to be decided once we have a clearer picture of what differentiates the two.

Case (3) should also count as testimony, though, like in Case (1), whether it is testimony-that or testimony-how is ambiguous. On the one hand, it might qualify as testimony-that, if A means to describe to B how A (personally) φs: A is simply offering testimony as to the steps she takes to φ. On the other hand, if A means to describe to B how *one* φs, using herself as an example, it seems that this case might count as testimony-how. This is especially true if A is counting on B to interpret her testimony as such.

On the face of it, Case (5) is slightly more unusual, but as long as we accept that testimony need not be spoken, but that non-verbal gestures carrying information might constitute the relevant kind of act of communication, then we find that Case (5) is testimony-how. Imagine that you (B) ask me (A), "How do you knit a C4L stitch?" and I reply, "Watch!" and take your knitting needles and demonstrate,

slowly, how to knit a C4L stitch. You observe intently, and after my demonstration, can flawlessly execute a C4L stitch, which turns out to be much simpler than you expected. You now know how to knit a C4L stitch precisely because I have instructed you, through demonstration, how to do so.

Let us also consider the more peculiar cases of Hawley's, especially those where B comes to learn how to φ from A, even though A might not have intended to instruct B. In particular, let us consider cases (4) and (6):

4. B overhears A talking to someone else [C] about how to φ (or about how A φs)
6. B observes A φ-ing and imitates A.

Here is an example of Case (4): Suppose an employee at a hardware store is instructing a customer in how to stain a deck. The employee says to the customer, "First, sand the wood so that it's ready to absorb the stain. Stir the pot of stain and then apply it to the edges and corners of the deck." Meanwhile, I am standing a few metres away, looking at paint chips for my kitchen, and overhear the employee's instructions. Though I was not in need of instructions on how to stain a deck, I now know the first few steps I would need to complete if I decided that the deck needed a facelift. In this case, I (B) have overheard the employee (A) talking to the customer (C) about how to φ. My intuition is that this case ought to count as testimony, but I shall pause for a moment here to consider Case (6) before discussing them both.

Case (6) is importantly different from the others we have considered so far. Here is an example: Suppose I am with my friends and we are dancing the Macarena, which comprises a fixed, repeated sequence of simple dance moves. You, unfamiliar with the dance, see us dancing and copy our motions from a distance. The sequence is simple enough that after a few repetitions of the dance, you memorise the sequence and can dance fluently. Much later, you hear the song playing, and successfully dance the Macarena by yourself. You now know how to dance the Macarena, and you know how to do so precisely because I (and my

friends) knew how and you observed us dancing. In this case, you (B) observed me (A) φ-ing, and imitated me. Your knowledge of how to dance the Macarena counterfactually depends on mine. And yet, intuitively, this case should not count as testimony.

What is different between the two cases? Both involve B learning how to φ from A even though A is not interacting directly with B. Why is it that one counts as testimony and the other does not? The answer lies in Lackey's formulation, in which *a*, an act of communication amounting to "the intentional expression of communicable content," (2008, 28) is a necessary condition for testimony.

In Case (4), the hardware store employee performs this act of communication by intentionally instructing the customer (and, inadvertently, me) in how to stain a deck. The employee performs an act of communication, and in doing so, reasonably intends to convey information, and that act of communication serves as a source of belief for both the customer and me. The instructions therefore count as testimony. In Case (6), however, there is no such intentional expression of communicable content. Thus, even though my dancing acts as a source of belief for you, and you come to know how to dance the Macarena, this is not an instance of testimony.

Adapting Lackey's view for testimony-how can account for a diverse set of cases in which B learns how to φ from A. It can rule out those cases that ought not to count as genuine testimony, while providing an account of those that should. Moreover, the close connection between the accounts of testimony-how and testimony-that will facilitate future research seeking to expand the notion of testimony to include testimony-how. Let us explore this idea in more detail in the next section.

2.7.2 Similarities between testimony-how and testimony-that

It is also reassuring to note that testimony-how, as I have described it, bears a number of similarities to testimony-that. I offer three examples of these similarities

below. First, testifying does not entail knowing; second, testimony can be demonstrative or verbal (Audi 2013); and third, there might be some version of the reductionist/antireductionist debate applicable to testimony-how and testimony-that alike. That testimony-how and testimony-that are similar is not an argument that they are the same; nevertheless, the family resemblance suggests that the two are related in an important way.

In Section 2.2.2, I accepted that one need not know that *p* in order to testify that *p*. The analogue is true in testifying-how: one need not know how to φ in order to testify how to φ. Consider this example from Hawley (2010): suppose you tell me that you can teach me how to carve a tomato rose, but unbeknownst to me, you do not actually know how to carve a tomato rose. Instead, you hack randomly at a tomato and luckily, it turns out to resemble a rose. I copy you, and it turns out that if I do what you did, I wind up with a tomato rose, too. Suppose I commit this method to memory and can now reliably correctly carve a tomato rose whenever I wish. Assuming there really is such a thing as testimonial knowledge-how, do I now know how to carve a tomato rose on the basis of your testimony? It would seem so, and this is hardly surprising. In both testimony-how and testimony-that, an agent can come to know something on the basis of a testifier's say-so, even when that testifier does not know that thing herself. One may, of course, be concerned about the possibility of Gettiered knowledge-how (Stanley and Willlamson 2001; *pace* Poston 2009), or concerned about epistemic luck (Shanton 2011), but the concern is *mutatis mutandis* the same as in the testimony-that case, and so provides no special worry here for testimony-how as a species of testimony.

Another similarity is that both testimony-that and testimony-how can take either verbal or demonstrative forms. There is an analogy between the following kinds of cases:

A asks B where the door is, and B points toward the door.
A asks B how to ride a bike, and B gets on a bike and shows A how to ride a bike.

And between these cases:

A asks B where the door is, and B says, "Just around the corner."
A asks B how to ride a bike, and B provides some detailed explanation of how to ride a bike.

In any case, B has learned something from A on the basis of some expression of A's knowledge. This accords with what Hawley says about the myriad ways B can learn how to φ from A, notably Case (5), which I discussed in Section 2.7.1. Moreover, there are similar considerations to be made about the significance of showing vs. telling; see, for instance, Habgood-Coote (2018b) on the relationship between the knowledge norm of assertion and the analogous knowledge-how norm of showing.

Finally, there is probably some version of the reductionist/anti-reductionist debate applicable to testimony-how. Reductionism about testimony is the view that testimony is not a fundamental source of knowledge, while anti-reductionism is the opposite. It will depend on what we think knowledge-how really is, especially whether it involves justification, or whether it is instead subject to success conditions (Hawley 2003; Stanley 2011). More work needs to be done to determine the role that justification plays in knowledge-how, especially if we take knowledge-how to be the kind of thing communicable via testimony. I remain neutral on exactly what the conditions for justification are for testimonially-based knowledge-how, but suggest that epistemologists ought to look at testifier reliability and trustworthiness in the first instance (see Hawley 2003; Peet forthcoming, esp. HONEST BOMB case).

Again, I wish to emphasise that the similarities alone are not conclusive proof that testimony-how and testimony-that just are the same thing. Rather, the idea is that if the two were sufficiently dissimilar, we would be unlikely to find analogous

concerns arising for each. These analogues, then, act as evidence that testimony-how and testimony-that are similar in some relevant, important way.

2.8. The memory-testimony analogy revisited

All of this has a follow-on implication that this final section is devoted to exploring. In Chapter 1, I argued that memory and testimony are analogous; my argument pertained, however, only to declarative memory and (what I am now calling) testimony-that. In the initial analogy, it really did make sense to talk about the analogy in terms of propositions, or instilling beliefs that *p* in either another person or in one's future self. In this chapter, however, I have argued that not all testimony is propositional. But this should not turn out to be a problem for us, because we already know that there is such a thing as procedural memory, or memory-how.

If we think of memory as the testimony of the past self, as Dummett (1994) did, we might say, crudely, that while declarative remembering is receiving testimony from the past self, procedural remembering is receiving testimony-*how* from, or being instructed by, the past self. This characterisation may not map onto the subjective phenomenal character of remembering-how, but that alone should not be reason to reject this hypothesis. Consider that the constructive nature of declarative memory, for example, does not map onto the subjective phenomenal character of remembering-that, but this does not cause us to reject the claim that declarative memory is constructive. In the declarative and procedural cases alike, we must hold in mind that memory at once *feels* like accessing an archive *and* at the cognitive level is a constructive process.

Procedural memory is often taken to be synonymous with non-declarative memory, but there are other types of non-declarative memory aside from procedural memory (Michaelian 2011a; Squire 1992; 2004; Squire and Zola-Morgan 1988; Tulving 2007; Werning and Cheng 2017). One might therefore think that if there is to be an analogy between the memory and testimony, there must be testimonial

counterpart for every kind of memory, including each subspecies of non-declarative memory. As an initial response, there may well turn out to be virtually countless ways that we convey information to others. For a start, Hawley (2010) names those I addressed in Section 2.3 and again in Section 2.7, but by her own admission, the list is far from exhaustive. There might be stronger connections and correspondences between certain kinds of non-declarative memory and certain kinds of testimony-how. I set this aside as a project for future research, but I hope that the line of inquiry I have opened here can serve as a springboard into investigating these connections.

2.9. Conclusion

Here, I wish to reconnect the points I have made about testimony to my earlier points about memory. I suggest tentatively that, to the degree that we might think of memory as being akin to the testimony of the past self, we might think of procedural memory as being akin to the testimony-*how* of the past self. In much the same way that we learn from others how to do things, and to the degree that that learning rests on our instructors' knowing what they are telling us to do, we might be able to perform certain actions precisely because our past selves instruct us in how to do so. I leave this only as a tentative suggestion for now, but suspect that an investigation into the psychology of procedural memory will reveal interesting connections between learning-*how* from others and from ourselves.

There are two ways to account for testimony-how. One is to develop a specific account of testimony-how; the other is to develop an account of testimony that accommodates both testimony-how and testimony-that. I favour the former option, but it is my hope—though it may be overly optimistic—that developing a specific account of testimony-how will serve as a first step to developing a general account of testimony that does not privilege testimony-that over testimony-how, or vice versa. By making explicit the nature of testimony-how, I hope that we can begin to move

toward a broader understanding of testimony that more accurately captures what is really going on when we learn from others.[12]

[12] I am grateful to Hannah Clark-Younger for helpful comments on this chapter, and to James McKeahnie for proofreading.

CHAPTER 3:
COLLECTIVE CONFABULATION

3.1. Introduction

In Chapter 1, I argued that memory and testimony were analogous; I showed that epistemologists should think about the two in the same way. Still, this characterisation does not go far enough. While the two are epistemologically analogous, thinking of them as *merely* analogous will inevitably prevent us from noticing that the two interact in several ways during ordinary life. I alluded to this in Chapter 1 when I pointed out that testimonial exchanges depend on the "informational richness" (Kenyon 2013) of the context in which they occur. As I claimed above, mindreading involves the attribution of mental states to another person, including their desires, motives, affective states, and background knowledge they already possess. Some of this information will come from the memory of the testifier; if, for example, I know you to be sensitive about a particular topic, I might tailor or truncate my testimony so that it is more palatable to you. In these ways (among others), memory can influence testimony.

The claim that testimony influences memory, however, is a harder pill to swallow. Surely, the folk intuition goes, my memories are mostly intact and immutable (aside from forgetting over time), and any testimony I receive about an event will only serve to jog my memory for things I have forgotten, but knew at one point. The curious case of the Mandela Effect,[1] presented in this chapter, suggests otherwise. And as we shall go on to see in Chapter 4, there are far more toxic ways for testimony to be integrated into memory. For now, though, let us consider the Mandela Effect.

[1] The term "Mandela Effect," as far as I can tell, was first used by self-described "paranormal consultant" Fiona Broome when she noticed the paradigm Mandela case (Broome 2009).

In recent years, a growing number of internet users have flocked to online fora such as Reddit to discuss *Shazaam*,[2] a 1990s film starring the American comedian Sinbad as an incompetent genie who grants wishes to two children. Reddit users, or Redditors, reminisce about what the cover of the VHS looked like, and recall a scene where candy rains from the sky (u/DonnaGail 2017)[3]. Some report fond memories[4] of quoting catchphrases from the movie with their siblings (Tait 2016), and one Redditor recalls having to watch the movie repeatedly to inspect the tape for defects as part of his job working in a video rental store (u/EpicJourneyMan 2016). None of this is unusual for a beloved children's movie, but *Shazaam* is unlike other movies in one crucial respect: it does not exist.[5]

Proposed explanations for the widespread memory beliefs about *Shazaam*—such as the suggestion that Redditors are thinking of the 1996 film *Kazaam*, starring basketball player Shaquille O'Neal as a genie—are often met with resistance and denial. Some claim to remember both *Shazaam* and *Kazaam*: one Redditor remembers deciding not to see *Kazaam* because it looked like an imitation of *Shazaam* (Tait 2016), while another remembers ordering two copies of *Shazaam* but only one of *Kazaam* for his video store (Tait 2016). Even though Sinbad himself has repeatedly denied having starred in such a film, some Redditors are so confident that *Shazaam* is real,

[2] Variously spelled "Shazam," "Shazzam," and "Shaazam."

[3] This citation style will seem foreign to readers unfamiliar with Reddit, but since the topic at issue requires frequent references to Reddit, here is an explanation: the prefix "u/" denotes that the name in question belongs to a user (rather than a thread or subreddit), and is followed by the username of the author of the relevant post. The prefix "r/" denotes a subreddit and is followed by the name of the subreddit.

[4] Where no confusion will result, I use the term "memory" and its cognates in a non-factive way. To call the representations in question "ME-memories" does not entail that the representations be veridical, but readers who insist on the factivity of "memory" can substitute "ME-memories" with "ME–apparent memories."

[5] To complicate matters further, in 2019 during the writing of this chapter, a real film titled *Shazam!* was released, directed by David Sandberg and starring Zachary Levi. This film is not the one at the centre of the Mandela Effect controversy, so can be safely disregarded for the purposes of this discussion.

they have offered cash rewards for proof of its existence (Tait 2016). Needless to say, all searches for proof have come up empty-handed.

As bizarre as it might seem that hundreds of people who have never met might remember a film that does not exist, it turns out that this case is anything but isolated. Some Redditors claim to remember that Nelson Mandela died in prison in the 1980s, though he was released from prison in 1990 and died in 2013 after a high-profile political career. But the belief that he died in the 1980s is not a matter of making a simple mistake about the facts, nor is it to remember correctly a falsehood that one was taught: the belief is supported by episodic memories, such as watching the news reports on television and discussing the event with family members or colleagues (Tait 2016).

The Mandela case serves as the paradigm case of a phenomenon known as the "Mandela Effect," wherein individual subjects who have not met offline develop highly similar memories of events that never occurred. Further examples of the effect are numerous and varied: some people remember watching evangelist Billy Graham's funeral on TV well before his actual death in 2018 (Broome 2013), that the *Monopoly* mascot Rich Uncle Pennybags as wearing a monocle (u/TimmehTheShpee 2018), Mother Teresa's canonisation by Pope John Paul II in the 1990s while she was still alive (u/ThadeusOfNazareth 2016), Leonardo DiCaprio's acceptance speech for the 1998 Academy Award for Best Actor for *Titanic* (Broome 2016), and so on. The memories in question are normally not intrinsically implausible: it is entirely possible that Sinbad could have starred in a movie called *Shazaam*, that Mandela could have died in prison in the 1980s, that DiCaprio won Best Actor for *Titanic*, and so on. Yet despite the subjects' inability to provide any nonmnemic evidence that what they report is actually true, subjects retain a high degree of confidence in the veridicality of their memories.

Discussions of the effect on Reddit have spawned the most detailed explanations available, ranging from the fanciful—that we are sliding between

alternate realities (u/AscendedMinds 2017)—to the more plausible but uninformative—"our memory isn't as good as we thought" (u/Jhoobie 2017). The superficiality of the latter kind of explanation leaves something to be desired, so I offer this chapter as an alternative.

Clearly, the Mandela Effect poses interesting questions for several branches of memory studies, but—perhaps owing in part to the novelty of the phenomenon—as far as I can tell, it has received very little scholarly attention. Furthermore, although both collective memory and memory errors are gaining popularity in the philosophy of memory, to my knowledge, there has been no explicit discussion about the nature of collective memory errors, no serious attempt to develop a taxonomy of collective memory errors, and no comparison between individual and collective memory errors.[6] Space prevents me from delving too much into these topics, but I intend to broach the subject by focusing on the unusual case presented by the memory representations that the Mandela Effect comprises (henceforth, "ME-memories").

This discussion in this chapter unfolds against the background of two ongoing debates in the philosophy of memory: first, the debate over the ontology of collective memory, including whether collective memory is ontologically distinct from individual memory, and second, the debate over the ontology of individual memory errors and how to distinguish them. I argue that ME-memories are (i) collective, since the representations that constitute ME-memories are direct results of the interaction of multiple individuals; (ii) matters of memory, since subjects make claims about the past on the basis of what they seem to remember; and (iii) errors, since subjects remain convinced of the truth of their memories, despite being aware that the overwhelming majority of people do not share their memories and despite the absence of nonmnemic evidence of the relevant events. If we take the notion of

[6] One exception is Tanesini's (2018) discussion of collective forgetting. It is not clear, however, that forgetting is properly understood as a memory error (Michaelian 2011b); even if it is, then it is an error of omission, while collective confabulation is an error of commission.

collective memory seriously, we should be prepared to take the notion of collective memory error seriously; combined with claims (i) through (iii), we reach the conclusion that ME-memories are instances of collective confabulation. This is the thesis I defend in this chapter.

There is one problem worth addressing at the outset: it is not clear that ME-memories are a homogeneous group. In addition to the *Shazaam* and Mandela cases, other examples are the cases of remembering the children's book series title as *Berenstein Bears* (it has always been *Berenstain Bears*), and remembering that cartoon character Carmen Sandiego had a yellow trenchcoat (it has always been red). These cases are importantly different from either the Mandela or *Shazaam* cases: namely, for reasons we shall see in Section 3.5, they are not necessarily collective in any robust sense, since people could independently misremember the title's spelling or the trenchcoat's colour. Nevertheless, Redditors classify cases like these as instances of the Mandela Effect. Though I lack the space to discuss such cases at length, I will say that it is likely that the Mandela Effect encompasses several different kinds of memory phenomena. Some may be genuinely collective while others may not; some may be matters of misremembering, some may be matters of confabulation, and still others may be conspiracy theories masquerading as memory reports. The heterogeneity of the category should be unsurprising—after all, informal Internet fora are not known for insisting on conceptual clarity—but taxonomising the category is not the main project in this chapter. For now, it will have to be enough to say that at least some ME-memories are collective confabulations, and to focus on how these cases come about. I remain agnostic on the rest.

Defending the claim that (some) ME-memories are collective confabulations requires showing that ME-memories are (i) collective, and (ii) confabulatory. In Section 3.2 of the chapter, I provide an overview of two prominent accounts of confabulation: reliability accounts and causal accounts. In Section 3.3, I argue that ME-memories are genuinely confabulatory under reliability accounts of memory

(Michaelian 2016a). However, the reliability account remains controversial. Causal theories dominate the landscape in the philosophy of memory, so I also show that one need not endorse reliability accounts to characterise ME-memories as collective confabulation. I do so by offering a modified version of my initial account that is consistent with causal theorists' characterisation of confabulation. In Section 3.4, I discuss accounts of collectivity. Finally, in Section 3.5, I show that ME-memories are genuinely collective, regardless of the account of confabulation one endorses.

3.2. Confabulation

"Confabulation" is a broad term including diverse phenomena, including both mnemic and non-mnemic confabulation. Non-mnemic confabulation can occur when people are asked to explain their attitudes or choices: despite being unaware of their reasons, subjects may nevertheless offer a sincere (but often incorrect) explanation (Bortolotti and Cox 2009), such as explaining why they had chosen the pair of stockings they did among a set of four identical pairs (Nisbett and Wilson 1977). In this chapter, rather than non-mnemic confabulation, I focus on mnemic confabulation: one kind of case in which the memory process produces something other than genuine memory.

I return to the details of confabulation shortly, but understanding the production of these non-genuine memories requires first understanding the production of genuine memories. Memory is now known to be a constructive process, wherein representations of events are not discretely stored, but reconstructed anew at the point of recall. Though several different accounts of the nature of constructive memory have been advanced, all converge on the central point that remembering is not the reproduction of a singular, discrete past event, but an act of reconstruction integrating information originating in several sources (De Brigard 2014; Michaelian 2011c; Robins 2016a; Schacter and Addis 2007). To use Bartlett's words, "Remembering is not the re-excitation of innumerable fixed, lifeless and fragmentary traces. It is an imaginative reconstruction, or construction, built out

of the relation of our attitude towards a whole active mass of organised past reactions or experience" (Bartlett 1932, 213). It is the reconstructive nature of memory that gives rise to confabulation and misremembering, whether they come about from reconstruction errors or filling in the blanks in the absence of relevant information.

Examples of misremembering and confabulation will prove helpful for the present discussion. Misremembering is a common error, involving the introduction of some inaccurate details into an otherwise accurate representation. Brewer and Treyens (1981), for example, conducted an experiment in which participants were exposed to a typical office scene. When those participants were later asked to recall what they had seen, many reported having seen a stapler, even though no stapler had been present. This is a characteristic example of misremembering: participants were able to represent a stapler *because* they had accurately represented the rest of the scene (Robins 2016a). By contrast, confabulations—what Schnider (2018) calls "normal false memory,"[7] or confabulations produced by subjects without memory disorders, as opposed to clinical confabulations—are more egregious errors. Loftus (1997b) showed how easily confabulations can be provoked in a study in which participants were asked to describe four childhood stories, three of which had actually happened and one of which—about being lost in a shopping mall at about age 5—was invented. In later interviews, when asked about the false event, 25% of participants reported fully or partially remembering the event (Loftus 1997b). Mnemic confabulations, loosely described, are apparent memories for events that the subject did not experience.

Roughly speaking, there are four families of accounts of mnemic confabulation: false belief, epistemic, reliability, and causal. Each provides a different set of conditions for confabulation, but a representation might

[7] Schnider (2018) is concerned with falsidical confabulation, but I use an account that allows for veridical confabulation.

simultaneously meet different accounts' criteria. For this reason, it may turn out that different accounts agree that a given representation is a confabulation while disagreeing about why that is so. I consider each type of account in turn.

3.2.1 False belief accounts

False belief accounts, as far as I can tell, have not received much uptake in philosophical literature, and are mostly confined to psychology. A number of different formulations have been proposed (e.g., Berlyne 1972; Feinberg 2001; Talland 1961; 1965; see Hirstein (2005), Chapter 8 for a summary), but all tend to agree that confabulations are sincere but false memory beliefs.[8] The main problem with false belief accounts is that they are simultaneously too broad, in that nearly any false belief qualifies as a confabulation, and too narrow, in that a confabulation that happens to be true does not qualify as a confabulation. The narrowness should concern us; as Hirstein points out, "A patient who gets a question right after supplying wrong answers to the previous six has not miraculously stopped confabulating" (2005, 199). The falsity condition alone is therefore inadequate to distinguish confabulations from non-confabulations, especially since we want to allow for the possibility of veridical confabulation. For brevity's sake, I end my discussion of false belief accounts here; nevertheless, it is worth noting that even if one were to endorse false belief accounts, ME-memories—such as recalling a news report about Mandela's death in the 1980s—would qualify as confabulations simply because they are false.

3.2.2 The epistemic account

Where false belief accounts posit falsity as a necessary condition for confabulation, for epistemic accounts, falsity is neither necessary nor sufficient for confabulation: what is more important is the fact that the belief is ill-grounded (unjustified), that the

[8] In fact, several accounts define confabulations as false statements or utterances, not mere beliefs, which makes the definition more restrictive still: such characterisations rule out the possibility of unstated but believed confabulations.

ill-groundedness of the belief is unknown to the subject, and that the subject should know that the belief is ill-grounded (Hirstein 2005). Note that the expectation that the subject *should* know that the belief is ill-grounded implies that she *can* know that her thought is ill-grounded, and yet because subjects are frequently not in a position to access the reasons for the ill-groundedness of their beliefs, it is not clear that confabulation is best defined in terms of ill-groundedness (Bortolotti and Cox 2009). If a subject lacks access to the reasons for her belief's ill-groundedness, then the representation would not qualify as a confabulation, and yet this seems unsatisfactory. Yet, the epistemic account is at a loss to explain such representations, and so, like false belief accounts, epistemic accounts are impoverished. I suspect, however, that the epistemic account—its shortcomings notwithstanding—would classify ME-memories as confabulations because they are ill-grounded beliefs whose ill-groundedness should be available to the subjects; i.e., subjects gain undefeated defeaters for their beliefs from external evidence, such as the testimony of others. Yet, subjects are culpably unresponsive to those undefeated defeaters. Thus, the representations in question qualify as confabulations.

3.2.3 Causal and reliability accounts

The false belief account tells an unsatisfactory story about confabulation. The case against the epistemic account is much less decisive; nevertheless, there are too many unanswered questions raised by the epistemic account for it to be useful in this analysis. For these reasons, I end my discussion of those accounts here and instead focus on reliability and causal accounts. Causal accounts dominate the theoretical landscape in philosophy of memory, but because causal and reliability accounts of confabulation have evolved in conversation with each other, I consider them side-by-side.

I begin with Robins' (2016a) causal account, which builds on the classic causal theory of memory (Bernecker 2010; Martin and Deutscher 1966). According to the causal theory, a subject remembers that p at some time t_2 if her present

representation that *p* bears an appropriate causal connection to her past representation at some previous time t_1 that *p* (Bernecker 2010). Robins develops a taxonomy of memory errors based on two conditions inspired by the causal theory: "retention of information from a particular past event and construction of an accurate representation of that event at the time of retrieval" (Robins 2016a, 445); since mnemic causation is normally understood as the retention of information, the retention condition is essentially a causal condition. Successful remembering, Robins argues, requires the satisfaction of both the retention and accuracy conditions: i.e., the information in the representation must have been retained from a particular past event, and the representation produced must be accurate with respect to the event. A failure to satisfy one or the other condition gives rise to various memory errors: failure to satisfy the accuracy condition brings about misremembering (inaccurate but retained representations), failure to satisfy the retention condition brings about relearning (accurate but not retained representations), and failure to satisfy either condition brings about confabulation. In other words, confabulation is a representation that is not produced from retained information, and is inaccurate with respect to the past.

On the face of it, the taxonomy is exhaustive, but it fails to distinguish between veridical relearning and veridical confabulation (since both would be accurate but not retained) (Bernecker 2017; Michaelian 2016b). Bernecker (2017) denies that relearning is a memory error at all and advances a modified causal account. Where Robins proposes that confabulation is a matter of a lack of *retention of information* from a particular past event, Bernecker proposes that confabulation is a matter of a lack of an *appropriate causal connection* to the past. The causal connection condition is slightly more permissive than the retention condition, since retention is just one way of several possible ways to satisfy the causal connection condition. Replacing the retention condition with an appropriate causal connection condition yields a taxonomy in which both successful remembering and misremembering bear

an appropriate causal connection, but the former is accurate while the latter is not. Denying that relearning is a memory error allows Bernecker to distinguish between veridical and falsidical[9] confabulation: both lack an appropriate causal connection to the past, but the former is accurate and the latter is not.[10]

Causal accounts, however, are not the only contenders in the philosophy of memory. Michaelian (2016b) advances a reliability account of memory errors based on the simulation theory of memory, according to which there is no essential difference between memory and imagination; to remember, as it turns out, is just to imagine the past (Michaelian 2016a).[11] Like its causal counterparts, the reliability account includes two conditions: reliability and accuracy, wherein successful remembering satisfies both conditions, misremembering satisfies the reliability condition only, veridical confabulation satisfies the accuracy condition only, and falsidical confabulation satisfies neither condition. Otherwise put, the reliability theorist replaces the causal theorist's causal connection condition with a reliability condition. Reliability theorists take a reliable system to be one that is functioning properly, so to claim that the system is unreliable is essentially to claim that the system is malfunctioning. The reliability theorist's argument in a nutshell, then, is this: a system is malfunctioning iff it is unreliable, confabulations are produced by unreliable systems, so confabulations (whether accurate or inaccurate) are produced by malfunctioning systems. In other words, the reliability theorist claims that a confabulation is a representation produced by an unreliable (or malfunctioning)

[9] "Falsidical" will be a new word to some readers. It is becoming more common in the memory literature to use the term to describe apparent memories with false content. The rationale, I believe, is to indicate that the content is false without committing oneself to the view that the apparent memory is thereby not a (real) memory, as is suggested by the term "false memory."

[10] For the remainder of the chapter, I rely on Bernecker's version of the causal account of confabulation, since it can account for both veridical and falsidical confabulation.

[11] Indeed, there are other interesting views of memory. Fernández (2018) proposes a functionalist theory of memory, according to which a memory state is the kind of state that tends to be (but is not necessarily) caused by a previous experience of the event represented. I leave the development of a functionalist account of confabulation as a task for future research.

memory system, where reliability does not require causation and is instead understood as a tendency to produce accurate representations.

So the question here is really this: is confabulation distinguished from successful remembering and misremembering by the lack of an appropriate causal connection to the past, as the causal theorist says? Or by unreliability in the memory system, as the reliability theorist says? The two accounts are on equal footing in that they are both able to distinguish among remembering, misremembering, and veridical and falsidical confabulation, so the debate will have to be decided on other grounds—grounds that I will not consider in the remainder of this chapter. For my purposes, it does not matter whether one endorses a reliability account or a causal account: as I show in Section 3.4, both accounts will characterise ME-memories as collective confabulations. The explanations of the mechanisms giving rise to ME-memories, however, will differ.

3.3. Are ME-memories confabulatory?

Arguing that ME-memories are confabulatory hinges on the claim that memory plays a central role in their formation. The astute reader may have noticed the striking resemblance between ME-memories and conspiracy theories, and would be justified in wondering whether ME-memories are really matters of memory at all. Both ME-rememberers and conspiracy theorists insist that reality is other than it appears and offer evidence—even if it is weak evidence—in support of their claims. For example, the organisation Architects & Engineers for 9/11 Truth (AE911Truth) disputes the conclusion that the impacts of the aircraft combined with the resultant fires were responsible for the collapse of the Twin Towers, claiming instead that the collapse was caused by a controlled explosive demolition (McDowell and AE911Truth Staff 2015). Similarly, ME-rememberers might dispute that Mandela was released from prison and did not die until 2013 by appealing to their memories of having read newspaper articles about his death in prison in the 1980s. The two

seem so similar as to appear to be the same thing. So, perhaps memory is a red herring; memory might play a central role in ME-memories, but it would be a mistake to think that they are worthy of philosophical attention as a unique phenomenon. Ultimately, the objection goes, ME-memories are ordinary conspiracy theories into which memory figures prominently.

But this objection is mistaken, for it obfuscates an important difference between ordinary conspiracy theories and ME-memories: while conspiracy theorists offer alternate *explanations* for observable events, ME-rememberers propose different events altogether. For example, AE911Truth does not dispute the observable events—namely, that the Twin Towers were struck by aircraft and subsequently collapsed—but they do dispute that the impact from the aircraft *caused* the collapse. They offer a different causal explanation for the collapse (namely, controlled demolition). By contrast, ME-rememberers *do* dispute events by, e.g., denying that Mandela died not in 2013 but in the 1980s and in prison; they offer different evidence for their claims about events. Granted, the evidence comes solely from their memories, and they are unable to provide any non-mnemic evidence to support their claims. For example, they might dispute Mandela's 2013 death (an event) by claiming to remember having read a newspaper article about his 1980s death (mnemic evidence) but searches for the newspaper articles themselves (non-mnemic evidence) yield nothing. So, ME-memories are not merely memory-flavoured conspiracy theories; the Mandela Effect is a unique phenomenon in its own right.

Interestingly, however, explanations of the Mandela Effect itself or explanations of the particular memories might themselves constitute conspiracy theories. For example, one might use the very existence of the Mandela Effect as evidence that parallel universes are real, or that we are living in a simulation, or that the government is systematically erasing records of the past and brainwashing our fellow citizens. Such a conspiracy theorist would not dispute that some remember Mandela's death in the 1980s and others remember his death in 2013, but would

attribute the dissonance to the existence of parallel universes. Similarly, one might claim that one reason so many people seem to remember *Shazaam* is because it really did exist, but the poor audience reception harmed Sinbad's career so severely that the studio erased all trace of its existence (u/squidink20 2018). In this case, there is no dispute over whether *Shazaam* actually existed, but instead a proposed explanation for its absence. Both of these examples should qualify as conspiracy theories, but note that the reason they qualify is that the *explanations* of the events differ; there is no dispute over the event itself. So, we can think of the Mandela Effect as involving two stages: at the first stage is the production of collective ME-memories themselves, while the second stage is a contingent downstream conspiracy theory to explain the Mandela Effect or the ME-memories themselves.

Having provided the necessary detail on what I take both collective memory and confabulation to be, and having defended the mnemicity of ME-memories, I now turn to the first of my two main aims: to show that ME-memories qualify as confabulations. I begin with the reliability account in Section 3.3.1, since it offers the most developed account of which representations qualify as confabulations and turns out to shed more light on the mechanisms by which those representations arise. In Section 3.3.2, I show that ME-memories meet the causal theorist's criteria for confabulation, so even if one rejects reliability accounts in favour of causal accounts, ME-memories still qualify as genuine confabulations. For now, however, I restrict my discussion to reliability accounts.

3.3.1 ME-Memories according to the reliability account

Recall that the reliability theorist classifies a representation as a mismemory when a properly functioning memory system produces an inaccurate representation, and as a confabulation when the system malfunctions. What this means for my purposes is that ME-memories will qualify as confabulations just in case they are produced by malfunctioning memory systems. The memory system in question is of a different kind than that in individual memory; where discussions about individual memory

are typically only interested in the function of the individual memory system, discussions about collective memory must attend not only to the individual-level processes, but the group-level processes responsible for producing group representations. Malfunctions at either level constitute malfunctions in the memory system, and therefore the reliability theorist would classify any resultant representations as confabulations (regardless of accuracy).

I propose that the collective memory systems, i.e., the collection of processes and individuals, that produce ME-memories are indeed malfunctioning, and that the malfunction lies at the group level. But if this is true, it only tells half the story, for ME-memories are produced, in part, by individual representations. If what I have proposed is right, then identifying the processes that give rise to ME-memories will require investigation of both the individual and group levels. In the following two subsections, I analyse these levels and identify the phenomena that precipitate ME-memories.

Individual-level factors giving rise to the Mandela Effect

The natural story to tell about collective confabulation is that it is simply a matter of confabulating together, but this story is mistaken. Under reliability accounts, confabulation is the product of a malfunctioning memory system. A malfunctioning system is one that functions unreliably, and so routinely fails to produce accurate representations; such a system does not behave according to its normal 'rules' and could thus produce a representation of nearly anything. Although it is possible in principle that multiple subjects could happen to confabulate the same thing, given the huge range of possible representations a malfunctioning memory system could produce, the odds of this happening in practice are vanishingly small (Hirstein 2005). But suppose that by chance, two subjects really did confabulate the same thing: this would still not explain ME-memories, for such a case would be an instance of merely shared and not genuinely collective remembering (more on this in Section 3.4), assuming that the subjects' interaction has not contributed new content

to the group representation. It is possible that the interaction among confabulators does contribute new content, but ME-memories are so widespread that this would fail to provide an adequate explanation of the appearance of ME-memories. So, if collective confabulation is not merely the result of interaction among confabulators, what explanation is left to the reliability theorist?

I argue in this subsection that collective confabulation involves misremembering at the individual level. To see how this might be the case, let us now shift focus from confabulation to misremembering, paradigmatically exemplified by the DRM Effect (Deese 1959; Roediger and McDermott 1995). In DRM-style experiments, subjects are presented with a list of related words such as *hospital*, *nurse*, *medication*, and *gurney*, and then when they are later asked to recall the words on the list, they report words that were not listed (such as *doctor*), but are thematically consistent with the presented words. Subjects are able to do this precisely because they have retained information[12] about related words that appeared on the list: the system reliably predicts words, getting some wrong and some right. Classifying the false memories as confabulations is tempting, but to do so would be to overlook the fact that the false memories are the products of a properly functioning memory system (Robins 2016a). If the memory system were malfunctioning, we would expect the reported but unlisted words to be thematically inconsistent with the words presented.[13]

[12] One need not appeal to the causal theory to offer this explanation of the DRM Effect. The reliability theorist can also explain the DRM Effect by appealing to retention, since the reliability theory does not deny that information is *ever* retained; it merely denies that information is *always* retained (Michaelian 2016a; 2016b).

[13] Here lies an objection. The version of the reliability theory upon which I rely takes misremembering to be the product of a properly functioning memory system that has produced an inaccurate output, but this should not be taken for granted. De Brigard (2014), for instance, argues that the function of the memory system is not merely remembering, but episodic hypothetical thinking (EHT). On this view, the system aims not at producing accurate representations of the past, but instead at producing plausible reconstructions of what *might have* been. If this is right, then accuracy is not a good indicator

At the individual level, the Mandela Effect appears to have a similar structure to the DRM Effect: subjects seem to remember an event that did not occur, but they do so because they have retained information about similar events that really did occur. For example, Sinbad really did have a film career, there really was a movie about a genie, and the words "shazaam" and "kazaam" are similar enough that sliding between the two is unsurprising; Nelson Mandela really was in prison and really did die (although not until 2013), and there were plenty of news reports about Mandela during the 1980s; DiCaprio has been nominated for several Academy Awards, although not for his otherwise critically-acclaimed performance in *Titanic* (which itself was also highly nominated). Examples like these align with what we already know about how a properly functioning memory system could seem to recall some false details precisely because it has retained information about related events. Rather than being due to a malfunctioning memory system, the individual representations that feed into the ME-memory are unlucky byproducts of properly functioning memory systems.

Misremembering is produced by a properly functioning memory system, and so the representations produced are guided by the memory system's function (namely, to get at the truth). So, if two subjects with properly functioning memory

---

of whether a system is functioning properly, since success at producing plausible hypothetical representations is not judged by accuracy. I suggest as a tentative response that discounting accuracy as a means to distinguish among genuine remembering and memory errors will prove to be a difficult move, as doing so renders the EHT view of memory unable to distinguish between genuine remembering or misremembering and confabulation: the content of the resultant representations might be inaccurate, but the inaccuracy cannot be attributed to a malfunction in the memory system (Robins 2017). If appealing to the cognitive process responsible for producing the representation fails as a means to distinguish errors, then the alternative is to appeal to the content of the representations and assess them for reasonableness. This strategy is undesirable, however, in that it will fail to catch mundane cases of confabulation (those that do not stand out by being sensational), and in that it "would invite a return to the unsavoury [psychiatric] practice of pathologizing deviance" (Robins 2017). In light of these considerations, we have good reason to preserve the notion of accuracy as an aim of the system, insofar as the process the system is executing is remembering (as opposed to imagination or counterfactual thinking, for instance). Therefore, it really is appropriate to take a properly functioning memory system to be one that reliably produces accurate representations of the past.

systems remember the same event, and if a properly functioning memory system is one that aims to produce accurate representations of the past, and if aiming at accuracy guides the representations produced, then it is relatively likely that two people could remember the same event in similarly inaccurate ways. An example will help to illustrate this point: two people might both seem to remember having seen a car collision resulting from one car running a stop sign rather than a yield sign. They remember the same inaccurate details, because their remembering is guided by aiming at truth. By contrast, a malfunctioning memory system is one that does not reliably—although it may occasionally—generate accurate representations. Because the resultant representations are not sensitive to the accuracy-aiming features of a properly functioning memory system, the likelihood of different people confabulating the same representation is down to chance—imagine that two people could seem to remember the same details about a collision that never occurred at all! The point here is that different subjects might spontaneously and independently misremember the same details as a result of memory's architecture. Let us consider how this works in the *Shazaam* case: is far more conceivable that two people could happen to remember the same inaccurate details about a movie that *does* exist (i.e. *Kazaam*), than it is that two people independently produced matching representations of a movie that does not exist. In both cases, the former is more likely, since a properly functioning memory system may nevertheless produce a few inaccuracies based on other remembered events. Although subjects are remembering inaccurately, the inaccuracies do not appear to result from malfunctioning memory systems; in other words, they are misremembering.

I claimed above that the reliability theorist would classify ME-memories as collective confabulations, i.e. collectively held representations produced by a malfunctioning memory system. So far, the story has only demonstrated how misremembering plays a role in collective confabulation; I have not identified any particular malfunction, and hence no confabulation. It is at this point that we must

consider the processes that unfold at the group level. Although I have argued that the Mandela Effect involves no malfunction at the individual level, in the next section I argue that there is a malfunction at the group level. It then follows that the collective ME-memory is a genuine confabulation: a representation produced by a malfunctioning memory system.

Group-level factors giving rise to the Mandela Effect

Let us return to the example of the car collision. Suppose, as above, that a witness misremembers a yield sign as a stop sign. Now, suppose that she later discusses the collision with other witnesses. When she mentions the stop sign, then if the other witnesses have remembered successfully, they might correct her by saying that it was a yield sign. But another possibility is that her interlocutors have taken her (inaccurate) testimony to heart and inadvertently incorporated the stop sign into their own representations: they have been influenced by the Misinformation Effect (Loftus 1997b; Loftus and Pickrell 1995; Loftus, Burns, and Miller 1978). If this is the case, the other witnesses might not correct our misrememberer, precisely because they are now also misrememberers with compatible representations. Whether the first witness is corrected or not, in both cases, the group representation tends toward convergence, although the convergence might be on an inaccurate representation rather than an accurate one.

But this does not appear to be what is happening in the Mandela Effect. In the above example, the group has converged on a representation of an event that *did* happen, but ME-memories are representations of events that *did not happen*. ME-memories are more akin to a scenario in which one witness to a car collision remembers a stop sign instead of a yield sign, another witness remembers a black car instead of a blue one, and still another witness remembers that the collision took place on a Thursday instead of a Wednesday. And of course, because people tend to misremember in similar ways, several witnesses may conceivably independently remember each of these details inaccurately. If these witnesses share their memory

of—or, in other words, testify about—the event in question, they may influence each other's representations such that the group representation they produce is so different from the actual event that it can hardly be said to represent that collision at all. In a case like this—and in the Mandela Effect—we are not seeing convergence on not just one, but multiple misremembered details. But why is it in these cases that we do not see a tendency toward accuracy? What features of interaction drive the formation of collective representations of events that never happened?

Let us pause here to recapitulate what I have said so far. I have claimed that the Mandela Effect is a matter of confabulation, which results from a malfunctioning memory system. I have also claimed that there is no malfunction at the individual level, and that the inaccuracies in question are mismemories, not confabulations. While for individual confabulation, the malfunction must lie in the individual memory system, the same is not true of group memory systems, since group memory systems, unlike individual memory systems, involve processing at both the individual and group level. I propose that the malfunction lies at the group level. Claiming that there is a malfunction at the group level—or indeed, claiming that there is a malfunction in any system—presupposes that there is some function that the system would normally execute, but has failed to do so. This forces us to answer the obvious question: what *is* the system's function? The short answer is convergence on true beliefs about the past. The longer answer will take some setup.

One initial worry that I should address before proceeding is that there may be no function at all; that it is an outright mistake to claim that groups have functions. And this may indeed be the case for most groups, but it is not obvious that the set of individuals that produce ME-memories are indeed groups of this kind. Some groups are merely nominal, such as the set of people on the same flight. Such a group lacks a function, although the members are engaged in the same activity (e.g., going to Beijing). But groups like these are constituted arbitrarily. By contrast, groups that are deliberately constituted—e.g., scientific teams, juries, committees, sports teams,

political parties—typically do have functions, such as winning a match or gaining economic and legislative control of a region. Groups like these seem to behave more as systems: where the actions are not just aggregations of individual actions, but involve group coordination and cooperation (Tollefsen 2015). Indeed, Staley goes so far as to insist that "collectives can only exist, and hence can only act, […] on the basis of shared aims" (2007, 322). Though this claim is very strong, it probably still points in the right direction: if not for a function (or intention or goal), it is unclear what would have motivated the members to assemble. Like sports teams, the systems that produce ME-memories ("ME-systems") are indeed deliberately constituted: the members find and engage with each other precisely because they share similar mismemories about the same events, not because they arbitrarily begin interacting. Otherwise put, the only reason these people even talk to each other is because they care about what happened. Thus, ME-systems appear to have some function.

If ME-systems have functions, they can malfunction. But there is an important difference between individual remembering and collective remembering. In the case of individual memory systems, it is easy to see how the function of the system can be understood in terms of the system's tendency to produce accurate representations (*pace* De Brigard 2014; see footnote 12 of this chapter). The same is not necessarily true of distributed memory systems, in which memory processing "spans not only the embodied brain and central nervous system, but also the environment with its social technological resources" (Barnier et al. 2008, 33). Although collective remembering can result in the production of accurate representations of the past, it also serves other purposes, such as promoting social bonds and reinforcing group membership and collective identity (Harris et al. 2014). If we imagine that the function of a distributed memory system is not to produce accurate representations of the past, but to fulfil one of these other social functions, our judgments about the reliability of the distributed memory system might change radically. Suppose a

married couple revises their memories about a past disagreement for the sake of harmony in the marriage. An example like this illustrates that a distributed memory system might fail to produce accurate representations of the past while promoting social bonds. If we take the latter, social function to be the function of the distributed memory system, then the failure to produce accurate representations should not figure into our judgments about whether the system is functioning reliably. So, the question is: are ME-systems more like individual memory systems, or more like distributed memory systems? What should we take the function to be: the production of accurate representations, or something else?

Suppose that ME-systems have a different function from individual memory systems. Perhaps ME-systems function to reinforce group membership; if this is the case, failing to produce accurate representations of the past would not act as sufficient justification for inferring that ME-systems are malfunctioning. If there is no malfunction, then ME-memories would not qualify as confabulatory according to the reliability account. The problem with the supposition, though, is that ME-systems exhibit different characteristics than typical distributed memory systems, and these characteristics invite caution in abandoning the notion of an epistemic function. For instance, one idiosyncrasy of ME-systems is that although the members could in principle come together offline, given the distribution of individual mismemories, this would be extremely unlikely: they would need to find each other offline by chance. ME-systems come together only online: their members must seek each other out, unlike the witnesses to a car collision who are brought together by chance. ME-memories are like DRM Effect memories in that they are products of properly functioning individual memory systems, but they are unlike DRM Effect memories in that any given ME-memory is unlikely to be widely shared, so for ME-rememberers to find each other requires that the members seek each other out. Indeed, in a typical case, the members have come to the online forum for the express purpose of finding out whether others remember the same past event, and the main

aim of discussion in the forum is to figure out whether the event in question really did occur. In other words, the ME-system has an explicitly epistemic function: to gain knowledge of the past. Many systems do not have genuinely epistemic aims, but it is not difficult to imagine ordinarily functionless interactions to have epistemic functions deliberately imposed upon them. For example, although the function of casual conversation is ordinarily to reinforce social bonds, some conversations have epistemic functions deliberately imposed upon them, as when someone asks a passer-by for directions. So, although it may sound odd to claim that this group interaction has a function, we can see that there are plenty of ordinary offline interactions that do seem to have functions imposed on them: for example, courtroom proceedings, academic conferences, or strategic planning meetings for a business. It seems that interaction within ME-systems could be of the same kind. Furthermore, the claim that the members are seeking to reinforce social bonds rings hollow here, since the members typically do not know each other offline and therefore have no pre-existing social bonds to reinforce. The systems in question would not even exist had the members not had the epistemic goal that motivated them to seek others who could confirm their beliefs.

Because ME-systems seem to have functions imposed on them, and because group formation happens as a direct result of the *epistemic* (not social) goals of the constituent members, we can treat ME-systems as having the function of enabling their members to get at the truth. In a sense, ME-systems are more like scientific communities, with epistemic goals including convergence on the truth (more on this in Section 3.5.2).[14] And although there may be other distributed memory systems that we consider to be properly functioning despite their failure to get at the truth, owing to the fact that they reliably fulfil some other function (such as reinforcing social bonds), it *is* the case that the success of ME-systems is indicated by their

[14] In fact, interactions within scientific communities are subject to the same kind of considerations that give rise to ME-memories, including a desire for agreement and group solidarity (Staley 2007).

reliability in getting at the truth. Yet, ME-systems systematically fail to perform this function, so we can infer that they are unreliable.

The problem with this inference, it might be objected, is that ME-systems typically only come together for the express purpose of discussing a single ME-memory, and exist for only a short time before dissolving. Because of their short lifespans, and because patterns are only established after repeated executions of the function, it is impossible to identify a pattern of reliability in a particular ME-system. In other words, the worry is that since ME-systems do not function long enough to establish a pattern, we can only observe a one-off inaccuracy rather than actual unreliability. Without unreliability in the system that produced a representation, the representation fails to qualify as a confabulation, or so the objection goes.

It might be true in many cases that a single system does not demonstrate unreliability, but there are a few things to be said in response to this objection. First, although many ME-systems are fleeting, some are persistent; i.e. the same group members might discuss several different ME-memories (in fact, some Redditors appear in multiple different but related Mandela Effect threads). A system of this kind might demonstrate a pattern of unreliability if it is systematically failing to get at the truth. But the second point is that even if we assume that all ME-systems are fleeting, we can generalise across ME-systems. Doing so shows that as a type, ME-systems systematically fail to get at the truth, and thus we can use information about the type to give us indirect insight into tokens of the type: namely, that individual ME-systems themselves are unreliable. And third, considering the patterns of interaction that propel the formation of ME-memories will enable us to infer that ME-systems are unreliable, since these patterns of interaction will routinely guide the system toward convergence, often at the expense of the representation's accuracy. (I return to this point in Section 3.5, where I discuss features of group interaction.)

In short, under reliability accounts, ME-memories qualify as confabulations. At the individual level, they are simple mismemories, but the routine failure of the epistemic group indicates a malfunction in the interaction among group members. Since confabulations are produced by malfunctioning memory systems, and since there is a malfunction in the collective memory system responsible for producing ME-memories, ME-memories are indeed confabulations.

This leads us to another objection. My opponent may wonder why ME-memories should be characterised as collective *confabulation* and not collective *misremembering*. Misremembering—paradigmatically exemplified by the DRM Effect—involves inaccuracy, but is produced by a properly functioning memory system (on the reliability view) or bears an appropriate causal connection to the past (on the causal view). It might seem as though the examples of ME-memories I have offered are quite close to the truth: there really is a movie about a genie granting wishes to children, and the titles *Kazaam* and *Shazaam* sound alike. It might appear as though the ME-memory is simply a matter of getting a few details wrong, and this, it seems, should not be enough to force us to characterise the representation as a confabulation. They seem more like inaccuracies introduced precisely because the memory system is working properly, just as is the case in the DRM Effect. It may seem as though ME-memories should therefore be understood as collective mismemories.

But to characterise them as such would be to miss a fundamental point in the reliability theorist's story: that mismemories arise when a properly functioning memory system produces an inaccurate representation. What is key here is that the misremembering/confabulation distinction does not map onto the accuracy/inaccuracy distinction; rather, it maps onto whether the memory system is functioning properly. As I argued earlier in this section, it appears that the memory system in question is malfunctioning; if so, then the system's outputs are confabulations, regardless of their accuracy with respect to the truth. At this point,

the reader may wonder why it is that the systems in question routinely produce "nearly true" representations; after all, this is not what we would expect from a malfunctioning memory system. The answer lies in the fact that the collective representation inherits information from the individual-level mismemories. Because the individual-level systems that produce these mismemories are not malfunctioning, and because the resultant representations are largely veridical, the collective representation will inherit both accurate and inaccurate features of events. Hence, a collective memory system could routinely produce "nearly true" representations even though it is malfunctioning.

Appealing to function will be unsatisfactory to the causal theorist, for whom the distinction between remembering/misremembering and confabulation depends not on function or malfunction, but on the existence of an appropriate causal connection to the past. While it is tempting to use a representation's inaccuracy as evidence of its confabulatoriness, the causal theorist denies that inaccuracy is what determines which representations count as confabulations. ME-memories qualify as mismemories if they have both an appropriate causal connection and are inaccurate. ME-memories do not meet this condition, however. As I will argue later in Section 3.3.2, ME-memories lack the appropriate causal connection. Therefore, the causal theorist cannot characterise individual ME-memories as mismemories, even though their proximity to the truth makes this verdict a tempting one. Instead, the causal theorist must characterise individual ME-memories as confabulations, but the story as to how collective ME-memories arise will be slightly different from the reliability theorist's.

3.3.2 ME-Memories according to the causal account

The reliability theorist can explain ME-memories by claiming that they result from individual-level misremembering combined with a group-level malfunction—in other words, ME-memories emerge when misrememberers interact. But this explanation fails the causal theorist, for whom what differentiates memory errors is

not whether the system that produced them is functioning properly, but whether they bear an appropriate causal link to a previous representation (Bernecker 2017). If ME-memories lack such a connection—and as I argue in this subsection, they do—they are confabulations. The purpose of this subsection is to tell a story about how the causal theorist can explain the appearance of ME-memories.

Let us return to a central question guiding what I have said so far in this subsection: what would the causal theorist need to say that the representations in question are confabulations? Importantly, the causal connection between the past event and the present representation must be inappropriate. The causal theorist who wants to call classify cases such as *Shazaam* as confabulation has two main options: first, to deny that there is an appropriate causal connection between the individual representations and the original event, and second, to deny that there is an appropriate connection between the individual representations and the group representation of the event. (You might think here that there is no such thing as a group representation at all. I address this notion in Section 3.4.)

The causal story begins in the same place as the reliability story, with inaccurate representations of events that really did happen. We start with representations of *Kazaam*, of news reports of Mandela in prison, and so on, but then details—such as the title or actor in the movie, or whether Mandela died or not—change, such that the representation is transformed into a mismemory. At this point, a misremembering subject who encounters evidence contrary to her apparent memory may seek assurances that her memory is intact, and upon searching online for what she remembers, she will find others who have reported similar mismemories. Like the reliability theorist, the causal theorist will appeal to memory's reconstructive nature to explain the preponderance of these mismemories (De Brigard 2014).

Once misrememberers congregate online, they share their own apparent memories about the event in question and integrate information from others'

testimony into their own representations, a phenomenon that aligns with Schnider's description of the mechanism that gives rise to everyday confabulation: "modification of the memory trace by the association of post-event information" (2018, 183). The ways that the topics are introduced influence both the content of and subjects' degree of confidence in the reported memories: if subjects are asked leading or suggestive questions by others, subjects' memories of certain events are distorted even before the memory is brought to mind, reproduced with strong or moderate conviction, and re-encoded (Schnider 2018, 182).

In the Mandela Effect, it appears that members' post-interaction representations depend much more heavily on the testimony of others than they do on any one past event. So, while the individual representations may have initially borne an appropriate causal connection to an actual past event—after all, there really was a movie about a genie, and the representation of that movie seems at least partially causally responsible for the present representation—and may even continue to bear a causal connection to the past, the notion that the link is appropriate is implausible. Although the initial representations bore an appropriate causal connection, this ceases to be the case once the causal connection has been mutated by the integration of post-event information in the form of testimony from other misrememberers. With the causal connection condition unmet, these cases will be classified as confabulations by the causal theorist.

But is the causal theorist to say that others' testimony renders the causal connection inappropriate? I think so. To see why, compare *Shazaam* to a car crash. Suppose several witnesses independently misremember different features of the car crash, and upon interaction, converge on a representation that turns out to be accurate. The question of the appropriateness of the causal connection is orthogonal to the question of the representation's accuracy, and so the verdict about the confabulatoriness of the car crash case should match the verdict about *Shazaam*, and ought to differ only as a function of the accuracy of the group representation. In

other words, if the causal theorist classifies the car crash as a case of genuine remembering, she should similarly classify *Shazaam* as a case of misremembering. If she classifies the car crash as a case of veridical confabulation, she should classify *Shazaam* as a case of falsidical confabulation.

What would the causal theorist say about the car crash case? It depends. The fact that the representation is accurate rules out the possibilities that the representation is collective misremembering and that it is collective falsidical confabulation, so the representation must be either genuine remembering or collective veridical confabulation. Which one it is will be determined by whether there is an appropriate causal connection between original event and group representation. So, to determine whether there is one, let us follow the representation right from the original event to the group representation it ultimately becomes.

First, there's the original event – say that what really happens is a white van runs a yield sign and hits a blue car at low speed. So then what? Suppose that three individual witnesses (prior to interaction) remember the following: Witness 1 remembers a white van running a stop sign and hitting a blue car at low speed, Witness 2 remembers a white van running a yield sign and hitting a black car at low speed, while Witness 3 remembers a white van running a yield sign and hitting a blue car at medium speed. In a sense, each witness has gotten the facts more or less right, but each has made an error attributable to ordinary constructive remembering, and which bears an appropriate causal connection (for example, Witness 2 remembers a black car *precisely because* she saw a blue car). Before interaction, each witness misremembers the collision.

Suppose now that the witnesses interact with each other. Suppose also that they are responsive to the fact that their errors are outnumbered (e.g., Witness 1, who incorrectly remembers a stop sign, is responsive to the fact that two others remember a yield sign, and updates his belief accordingly). In such a case, all the

individual-level errors will be corrected, such that the group representation is veridical: Witness 1's mismemory of the stop sign, Witness 2's memory of the black car, and Witness 3's memory of the higher speed of the collision are all corrected by the testimony of the other witnesses, which all trace back to the original event through what is clearly an appropriate causal connection. No problem here for the causal theorist: the collective representation is both accurate and appropriately causally linked to the past, so it qualifies as successful remembering.

But what if the witnesses interact in such a way that their errors survive interaction and make it into the collective representation? Suppose that during interaction, Witness 1 is especially insistent that the van ran a stop sign and not a yield sign, and that error makes it into the collective representation, such that the group representation is of a white van running a stop sign and hitting a blue car at low speed.

Here we meet a problem, and it comes down to the question of how to individuate events. If we individuate events narrowly, i.e. that a white van runs a *yield* sign and hits a blue car at low speed, this case will count as confabulation, because it was simply not the case that a white van ran a *stop* sign and hit a blue car. The question is whether the event and the representation are causally related; trivially, there is no causal connection to an event that never occurred. By virtue of the lack of causal connection, the representation qualifies as a confabulation.

Broader individuation is trickier: it depends on how many errors (and what kind) are tolerable. If we individuate the event more broadly—for example, we demarcate the relevant event as being a car collision—then it looks more like misremembering. The collision did happen, after all, but when the witnesses recall the event, they get some of the details wrong. So, there is a causal connection to a past event, i.e. the car collision, but the representation lacks accuracy. By Bernecker's lights, this qualifies as misremembering.

So, let us return to the question of *Shazaam*. What I have said so far depends on a fairly broad individuation of the past event in question, but the case for confabulation becomes that much stronger should we individuate the event more narrowly still. Suppose that the event in question is not whether there really was a movie about a genie, but instead whether there really was a movie about a genie, that was called *Shazaam* and starred Sinbad. The same integration of post-event information will unfold, but with this narrow individuation, the need to appeal to that integration to explain the erosion of the causal connection disappears: the event simply did not happen, so there is no causal connection to erode.

Still, it seems like a broad individuation of events is going to be the wrong avenue for the causal theorist to take. Witness 1 isn't insistent that *a white car ran a stop sign and hit a blue car at low speed*; most of those features are already agreed upon by the other witnesses. Witness 1 insists that a white car ran *a stop sign* and hit a blue car at low speed. So, it seems that this small aspect of the event is what is at stake here, and the causal history of that small aspect matters very much. The causal theorist cannot appeal to the whole representation's causal connection to the event when what is at stake amongst the rememberers is this narrow feature. In a sense, what the rememberers disagree about already presupposes a narrow individuation of events. So, the causal theorist will need to adopt a similarly narrow individuation.

If this is right, then what really matters is the causal connection of the stop sign to the past event, and trivially, there can be no causal connection to an event that never happened. Thus, since there is no causal connection, the causal theorist must conclude that the representation is a confabulation. The causal theorist might be perfectly happy for appropriate causal connections to make their way through testimony, but surely testimony about something that didn't happen won't qualify as an appropriate causal connection. So, to the extent that the collective representation depends on this specious testimony, it lacks the appropriate causal connection, and is therefore confabulatory.

But beyond an appeal to intuition, how are we do defend the claim that the causal chain is deviant? Consider the case of Kent, who experiences a car accident, recounts the story to his friend Gray, and then gets into another accident in which he sustains an injury causing loss of memory of the first accident (Martin and Deutscher 1966). Gray repeats the account Kent offered to him, and Kent comes to represent the collision once more, and later forgets that Gray has testified as to the details of the event. Here, Kent's representation of the car crash is veridical, so it cannot count as either falsidical confabulation or misremembering, since those are necessarily false. That leaves us with veridical confabulation or genuine remembering as the alternatives. Martin and Deutscher deny that Kent is remembering, so he must be confabulating, though veridically. Simply put, the representation lacks an appropriate causal connection, and is veridical, and is therefore veridical confabulation.

Whatever the verdict in Kent's case may be, it ought to match to the verdict in our own car crash example above. If the causal connection is lacking in Kent's scenario, then it is lacking in our own car crash example as well. Simply put, the car crash example is an instance of confabulation; the first iteration is veridical confabulation, the second falsidical. Recall that veridical confabulation and genuine remembering differ from each other in exactly the same way that falsidical confabulation and misremembering differ from each other (i.e. they are not matters of accuracy, but instead matters of whether the representation is appropriately causally connected to the event). Thus, if the second iteration of the car crash qualifies as falsidical confabulation, then so too should ME-memories.

Thus, whether the past event is individuated broadly or narrowly, the causal connection condition is unmet: either the subject is representing an event that simply did not occur, in which case there can be no causal connection, or the causal connection is specious, travelling through a tendrilous chain of mismemory and others' testimony through several subjects to a past event, such that the causal

connection can hardly be said to be appropriate. Either way, at the individual level, the causal theorist will characterise the representations in question as confabulations. What happens is that individuals form mismemories that are then distorted by testimony from other witnesses, producing representations whose deviant causal chain renders them confabulatory. That those confabulations come about as the result of interaction among group members means that the representations are genuinely collective, and thus collective confabulation.

In short, in either the reliability or causal case, the fact that ME-memories seem to be fairly close to the truth is inconsequential to determining whether they are confabulations as opposed to mismemories. The confabulatoriness of a representation is determined not by accuracy but by the reliability of the system that produced it (according to the reliability theorist) or by the representation's causal connection to the past (according to the causal theorist). Thus, while it is tempting to say that ME-memories are mismemories rather than confabulations, neither the reliability nor the causal theorist will arrive at such a conclusion; instead, both will classify ME-memories as genuinely confabulatory. This brings us halfway to the conclusion that the Mandela Effect is collective confabulation.

3.4. Collective Memory

Just as the purpose of Section 3.2 was to provide an overview of the main accounts of confabulation, the purpose of this section is to provide some background on what I take collective memory to be. What follows in this section will set the stage by providing the criteria that ME-memories must meet to qualify as a genuinely collective phenomenon.

Theiner (2013) argues that if a property of a system is to be considered emergent (*sensu* Wimsatt's (1986) mechanistic approach to emergence), it must be the case that interactions among the system's components affect that property (Michaelian and Sutton 2017). So, an operation performed by a system with a given

set of components that interact in a certain way will produce different results than a system with some components exchanged for others, or different interactions in place. So if the group members—in this case, individual rememberers—interact in cooperative or inhibitory ways, the output of the system will be affected. In other words, the fact that a given process is unfolding in a collective context has a unique and important impact on the output of that process. In a similar vein, Huebner (2014) argues that we can attribute collective mental states as long as the computations performed by the group through the exercise of whatever mental capacity we attribute to it are more complex than those performed by its members.

In both Wimsatt's/Theiner's and Huebner's approaches, the main point for my purposes is that collective mentality emerges through interaction among group members. Of course, not all group interaction is created equal, and the degrees of interactivity differ from one group to the next (Michaelian and Sutton 2017). In the case of remembering, we can identify two key processes in which we might expect to observe interaction: encoding (the point of committing a certain experience to memory) and retrieval (the point of accessing the encoded representation),[15] either of which can be parallel or interactive. A parallel process is one in which group members perform some task (e.g., encoding or retrieving) alongside one another but with no real interaction, and an interactive process is one in which group members interact while performing the task in ways that influence the output of the task (e.g., what is encoded or retrieved). This yields four possible combinations in the parallel/interactive encoding/retrieval framework: parallel encoding with parallel retrieval, interactive encoding with parallel retrieval, parallel encoding with interactive retrieval, and interactive encoding with interactive retrieval (Michaelian and Sutton 2017).

[15] This is a necessarily simplified view of the stages of remembering, but will suffice for my purposes.

The most robustly collective forms of emergence[16], interactive encoding and interactive retrieval, can be found in transactive memory systems (TMSs). In TMSs, the members' metacognitive knowledge and cooperative and inhibitory interactions are critical to the successful functioning of the system. But is it not merely that these interactions are critical to the successful functioning of the system; owing to these interactions, the computations performed by a TMS are more complex than those performed by its members (Huebner 2016). Both Theiner's and Huebner's claims appeal to the interactivity within TMSs to show that TMSs are capable of remembering content different from what any individual group member remembers. In other words, interaction produces a group representation that is quantitatively greater[17] than the representation produced by an individual; the interactions among components of the system influence the output of the system, which on either Theiner's or Huebner's views suggests that TMSs are genuinely collective. As we will see in the next section, ME-memories are likely not this robustly collective—I argue that they are matters of parallel encoding and interactive retrieval—but are collective nonetheless.

Collective memory should be distinguished from shared memory. For this chapter, I largely rely on Michaelian and Sutton's (2017) line on collectivity, according to which, while collective memories are shared representations resulting from interaction among group members, shared memory lacks the interactive

[16] I am not wedded to any particular technical understanding of emergence; the view I have in mind is fairly generic.

[17] In addition to quantitative forms of emergence in TMSs, there are qualitative forms of emergence: Harris et al. (2017; 2014), for instance, identify several kinds of emergence in collaborative remembering in long-married couples who "go episodic." These dyads exhibit the same type of quantitative emergence mentioned above, meaning that information that neither individual could recall alone becomes available due to interaction during retrieval. But they also exhibit qualitative forms of emergence, such as enrichment of the emotional features or the vivacity of what is remembered. Furthermore, members' interpretations of a given event might be transformed when they remember together. It seems clear, then, that remembering together is not just a matter of adding up what is remembered apart. I leave aside the topic of qualitative emergence in ME-systems, but offer the cursory observation that ME-system members' convictions in their representations increases with interaction, so it is plausible that ME-systems exhibit other qualitative forms of emergence.

component at either the encoding or retrieval stages (Michaelian and Sutton 2017). In other words, collective memory is a matter of parallel encoding coupled with parallel retrieval. To understand the difference, consider what it would be like to reminisce about a landmark football game with a friend who had attended it with you. You would likely remind each other of this pass or that goal, bringing up different details to each other that you had forgotten. Compare this scenario with one in which another spectator, someone you had never met and never would meet, were sitting on the other side of the stadium during the game. You and the unknown spectator might represent exactly the same features of the game—even simultaneously—but this match is obviously not a result of interaction between the two of you. This latter scenario is an example of shared remembering. Although your representation matches the unknown spectator's, there is no sense in which the memory is anything more than shared; interaction has not played any role in producing the representations.

I return to the distinction between collective and shared memory toward the end of this chapter, but it will not be too important for what follows next. Crucially, I take a collective phenomenon to be one in which interaction among group members influences the output of the process, such that the output is transformed from what would have resulted had any individual performed the process alone. In the next section, I use the claims I have advanced here to show that ME-memories are genuinely collective. I argue, in other words, that in-group interaction contributes to the formation of ME-memories.

3.5. Are ME-memories collective?

I claimed above that remembering is collective when new information emerges due to the interaction among group members (for example, new details, new qualitative information like emotional richness, or new understandings or interpretations of past events). In the case of the Mandela Effect, we are lucky to have a record of the

group interactions readily available for analysis: namely, the forum threads in which these ME-memories appear. It is clear upon reading these threads that members contribute new information that is readily adopted by other members, as when one Redditor writes:

> "*I read your synopsis and it's very close to what I remember.* I actually owned a copy of this movie that my mom bought from a video store because it was only like $1. One additional scene that I do remember involved a car and the kids wanting the genie to come with them somewhere but he couldn't sit in the car so he was riding on top of it like it was a flying carpet and they were like "No, you can't do that either! That's dangerous/someone will see you!" Cause he kept like almost hitting trees and things and sliding around (which never made sense cause if he's a genie couldn't he use magic? Idk [I don't know]. Lots of plot holes in this masterpiece) and people weren't supposed to see him or whatever. So then he disappeared and they were like "Where'd he go?" And they couldn't find him again until they got out at their destination and he was in the trunk and his body was all like twisted around weird and the kids thought it was so funny. *I was just curious to see if you remembered anything like that?*" (u/manafrmhvn 2018, emphasis added)

Another Redditor replies:

> "I don't want to inadvertently add to the "Mythos" surrounding this film by adding things that I am not 100% sure about, which is why I have never referred to the movie as "Shazzam" or any variation thereof for eaxample [*sic*] (it was a one word Title and the genie may have used it as a magic word but I can't say for sure that was the name of the movie).
>
> *I can say that yes, there was a whole segment of the film that involved Sinbad hiding and trying not to be seen, and the car scene sounds familiar* and I think had to do with the dad accidentally taking the bottle to work with him but I can't elaborate much more than that other than I think the dad nailed a presentation or meeting because the genie helped him without him knowing.
>
> I really think the movie was never finished being edited in post production and it was hurriedly released when the Rights to it changed hands to take advantage of Sinbad's popularity at the time.

> I actually wouldn't be surprised at all to find out the movie was originally filmed in 1989-90 before he was a big star...and another thing, might be nothing - but I could have swore [*sic*] the kids actually called him "Sinbad" in the movie...though I guess if it was "Shazam" it's pretty close phonetically." (u/EpicJourneyMan 2018, emphasis added)

These interactions illustrate how individuals (falsely, though likely sincerely) report familiarity with the information offered by others, and contribute their own memories or interpretations of the event. The individual narratives are woven together in such a way that ultimately, the group-level ME-memory is a composite of that information and is greater in detail than the representations produced by any one individual. In fact, Redditor u/shazaamthemovie compiled a list of 'known' information about *Shazaam* based on "stuff that multiple people from various sources remember" (u/shazaamthemovie 2017), including the suspected release date, starring actors, a description of the VHS cover, and details about particular scenes. Furthermore, as I claimed above, the interactions among ME-system members lead to the emergence of—at minimum—new details, even though these interactions are malfunctional. It is also possible that qualitative forms of emergence come about in ME-systems, but I leave that possibility aside for now. But if it is true that interaction leads to new information, then it follows that the group-level ME-memories are genuinely collective.

3.5.1 Group-level malfunction

My earlier explanation of the group-level malfunction (Section 3.3.1) was inspired by the simulation theory (Michaelian 2016a), but my explanation of the mechanisms that bring about this malfunction is compatible with the causal theory. Indeed, the causal theorist will likely need to appeal to these same mechanisms to explain how interaction among group members influences the resultant representation, and thus how individual mismemories are transformed into individual confabulations. The causal theorist need not characterise anything in the mechanistic story as a

malfunction, however, and—like the causal theory of memory itself—can remain agnostic about the notion of good functioning.

To understand how interaction shapes representations, consider a subject with a properly functioning memory system. The reconstructive nature of remembering might result in the subject's misremembering a certain event; ordinarily, this will not give rise to a particularly interesting collective phenomenon, since the subject might not relay the mismemory to anyone else. Suppose, however, that the subject *does* relay the memory in an offline environment: it is likely that her mismemory will be corrected by her (successfully remembering) peers. But her belief may be insensitive to correction, and if this is so, she may seek confirmation of the veracity of her memories elsewhere. It is possible that she might find such confirmation offline, but given that the distribution of mismemories is likely to be fairly scattered, her odds of encountering someone else with the same mismemory are low. Thus, even if she herself is unwilling to give up the mismemory, she is unlikely to convince others of its truth and it will remain a one-off inaccuracy: she will be the only one in her epistemic community who holds that mismemory.

But suppose our subject turns online to seek the confirmation that she is unable to find in her offline epistemic community. If enough other people with sufficiently similar mismemories have had their own offline failures to find confirmation of the veracity of their representations, they, too, may have turned to online discussion fora such as Reddit to seek such confirmation there. Since the members of such fora compose concentrated groups of misrememberers, the chances of in-group correction are significantly lower. The formation of the group of misrememberers triggers the operation of a variety of mechanisms, ultimately resulting in the reinforcement and enrichment of the individual subjects' mismemories. This, in turn, gives rise to a collective representation of the event.

3.5.2 Features of interaction

What exactly are the mechanisms at play here?[18] The first is one that I have already hinted at: the widespread phenomenon of online echo chambers,[19] in which "most available information conforms to pre-existing attitudes and biases" (Lewandowsky, Ecker, and Cook 2017, 359). In fact, the subreddit r/MandelaEffect allows users to filter posts by "skeptic" or "no skeptic" tags (u/Denominax 2017); conceivably, this facilitates the segregation of "believers" from skeptics. Even in other threads that do not explicitly forbid the expression of skeptical views, members who do express skepticism are strongly ostracised, as demonstrated by the response to a particular sceptical comment[20] on a thread titled "254 Confirmed Mandela Effects: List" (u/ezydown 2017). U/melossinglets challenges the comment with:

> "firstly,why dont you go to your precious google and look up the meaning of the word "skeptic"...then,once youve let that set in and marinate a little,see if you reckon that that definition correlates nicely with "a bunch of people coming in and simply telling everyone they disagree with that they are wrong and trotting out the exact same "cover-all" excuse for hundreds and hundreds of folk theyve never met in their life,basically making one huge assumption about all of their various experiences and painting them all with the same brush".....im not entirely sure it will.....but cool,whatever." (u/melossinglets 2017)

U/melossinglets' comment is characteristic of responses to skeptical views and illustrates one of the epistemic problems embedded in groups like this. First, there is a selection bias in that only people who share mismemories about certain events are

[18] I intend this subsection only as a first step toward identifying the mechanisms at play in producing ME-memories. It is not meant to serve as a definitive list.

[19] Here, one might argue that the echo chamber problem—i.e. the tendency to engage only with media that serve to confirm one's previously-held beliefs—is overstated (Dubois and Blank 2018; Garrett 2017). This may be the case, but the studies in question make this point with respect to politics. There may be some important differences between how echo chambers affect political discourse and how they affect collective remembering. Furthermore, my claim is not that echo chambers alone can explain the Mandela Effect, but that echo chambers exacerbate certain features of interaction—features that could, both in principle and in practice, be found offline as well as online—that are operative in the production of ME-memories.

[20] The original comment has been deleted, but the subsequent comments clearly indicate that the original comment expressed a skeptical view.

likely to find each other in the first place, so the balance of beliefs is tipped toward inaccuracy rather than accuracy. Those who deny these mismemories are forced out of the group, whether by formal penalty (such as a moderator ban on the user) or by social exclusion. Furthermore, in collectives of ideologically-aligned subjects, "as a form of 'identity self-defense,' individuals are unconsciously motivated to resist empirical assertions […] if those assertions run contrary to the dominant belief within their groups" (Kahan 2012, 408). In a ME-system, this heavily decreases the likelihood of correction and goes some way toward explaining the absence of a tendency toward convergence on the truth: simply put, members withhold dissent in favour of preserving group membership and solidarity. In a similar vein, we should bear in mind that groups aim not only at truth, but at agreement: members may act as though they believe things—i.e. they may *accept* aspects of propositions without *believing* them, so as to facilitate action (Tollefsen 2015), and therefore it might be the case that the group believes something that no individual member believes[21]. Speaking about scientific publications, Rehg and Staley (2008, 10) call this "*heterogeneous consensus*, in which a collaboration agrees to the publication of an evidence claim, while disagreeing on the premises offered in that publication as support for the claim." For instance, in a huge collaboration involving 450 (!) individual researchers, those who endorsed the findings of that collaboration overwhelmingly reported that the conclusion was "basically correct," but disagreed on other points pertaining to the conclusion (Staley 2007). Similarly, ME-system members are willing to endorse the general ME-memory even if certain aspects of it conflict with their own representations. Once a cohesive group has been established, members might be encouraged to continue to endorse their apparent memories for the sake of group stability and continuity. This is especially true for minority views—when one individual seems to be the only one who holds a certain belief,

[21] There is a rich and vast literature on exactly this topic, but space limitations prevent me from engaging with anything but a generic version of the claim. For more, see e.g., Hakli (2006; 2007), Tuomela (1992), and Wray (2001).

finding others who also hold that belief can galvanise it and make those believers that much less willing to renounce that belief.

Once the group is established, members discuss the details of the event in question. In fact, this is the express purpose of many such fora: the subreddit r/MandelaEffect, for instance, has a permanent post with the instructions, "Do you believe you've discovered a new Mandela Effect? Post it in the comments below to see if anyone else has experienced it too! Make sure you include why you think it could be a Mandela Effect and *as many details as possible so people can respond and discuss with what they remember*" (u/AutoModerator 2018, emphasis added). The entire point is to engage in collective remembering, evidenced by the straightforward invitation for members to imagine their own experiences and to contribute details to the group representation (u/DonnaGail 2017). Consider the following report from one Redditor:

> "Three of my coworkers and I were talking about a product we have, and the name sounded similar to Sinbad. We ended up discussing the actual movie called Sinbad, and then it led to discussing the comedian. I listened to this exact conversation go down between the three of them:
>
> 1: 'Yeah, he was in a movie! Umm ... he was like a genie or something.' 2: 'Oh yeah, I remember that. It was called .... Shazam? Oh wait, that was Shaq.' 3: 'No, no. That was Kazam. Different movie.' 2: 'Oh okay. Man, I haven't seen Shazam in so long.'
>
> I had no influence on the discussion. One sort of knows about the Mandela Effect, but I confirmed after this conversation that he had no idea that there was anything about the existence of Shazam." (u/Fae_Leaf 2018)

The contributed details cue other members to contribute what they remember, too, eliciting additional details from each other, and so on and so forth, gradually producing and refining a collective representation. This cross-cueing is particularly important, especially when the information being presented is false. If this is the case, we observe the Misinformation Effect: "the impairment in memory for the past

that arises after exposure to misleading information" (Loftus 2005). When participants present information about their own mismemories, this information—even if false—is incorporated into the memories of others.

Echo chambers, the prevention of expressions of skepticism, the invitation for individuals to offer their memories to the group, the misinformation effect: all of these are mechanisms that contribute to the formation of a collective representation of an event. It should also be noted, however, that in principle, none of what I have said is particularly unique to online environments. The onlineness of the interactions exacerbates the echo chamber problem, to be sure, but echo chambers of many kinds are found offline as well, since friend groups are often composed of like-minded people who already share viewpoints or attitudes, and in which discussing controversial issues rings of "preaching to the choir." And the other mechanisms I have identified will be exacerbated by the homogeneity of the ME-systems, but are still functional in offline environments. So, although the only cases of the Mandela Effect that I have observed have been online, there is no reason in principle that they could not emerge offline. Due to my point above about the offline distribution of mismemories, such emergence is unlikely, but if by chance several misrememberers were to find each other offline, we have no reason to think that a Mandela Effect-style interaction is impossible.

3.5.3 Large-scale or small-scale?

There is one final point to be made about collectivity and the Mandela Effect. Most of the features of interaction I have identified have pertained to small-scale groups, such as siblings, long-term friends, and intimate couples. In these cases, members interact directly with one another; it is this feature that enables the features of interaction I describe. By contrast, ME-systems seem to be comparatively much larger, and so one might wonder whether it is legitimate to apply concepts from research into small-scale groups to ME-systems, which are apparently large-scale. So the question here is this: if we are not willing to take the notion of large-scale

collective memory seriously, are we entitled to talk about collective confabulation, especially given that ME-systems seem to be so large-scale?

The answer is yes. Recall that what is necessary for a phenomenon to qualify as collective is that a group's constituent members interact in such a way that there is some degree of emergence. We see the most robustly collective kind of interaction in TMSs, where encoding and retrieval are both interactive. Though we do not see this kind of interaction in ME-systems, we still see parallel encoding and interactive retrieval, which, although weaker than the form TMSs exhibit, is sufficient for collectivity. The nature of an online forum enables direct interaction among members, such that the interaction is much more like the kind we would see in small offline groups; in other words, because members can interact directly with each other online, the group behaves less like a large-scale group and more like a small-scale group, even though the group may appear too large to function as such. The point is this: because ME-systems are sufficiently interactive—i.e., because the interaction shapes the representation produced—it is indeed legitimate to borrow features of interaction that have been identified in small-scale cases and apply them to the specific case of ME-systems.

3.6. Conclusion

I began this chapter by thinking that, if one takes seriously the notions of collective memory and memory errors, there must be such a thing as a collective memory error; nevertheless, there seems to be no discussion of such a thing in the literature. The relatively recent phenomenon known as the Mandela Effect serves as a fascinating example of the kind of thing we would expect if we were to imagine what a collective memory error might be. I have shown in this chapter that the Mandela Effect can be explained as collective confabulation, regardless of whether one endorses a reliability account or a causal account of confabulation. For the reliability theorist, the widespread misremembering of real events at the individual

level coupled with a malfunction at the group level produces a group-level confabulation. For the causal theorist, the group representation inherits the confabulatoriness of widespread confabulations at the individual level. In both cases, interaction at the group level promotes quantitative (and probably qualitative) emergence, such that the system remembers more than what any individual remembers. The phenomenon is thus genuinely collective.

I conclude with an exploratory note. As I said above, it is likely that the group representations that fall under the Mandela Effect umbrella are not a homogeneous group. Remembering *Berenstein Bears* instead of *Berenstain Bears*, or remembering that Carmen Sandiego had a yellow trenchcoat instead of a red one are cases that seem importantly different from fabricating an entire movie that never existed. One important difference between these and *Shazaam* are that they seem to be errors about simpler (i.e. more atomistic) things: there are only so many mistakes one can make when it comes to the spelling of "Berenstain," but as we have seen, the mistakes one can make about a film are seemingly endless. So, it seems much more likely that several individuals could independently produce the same representation without any interaction whatsoever. This may, therefore, be a case of shared misremembering: a matter of a reliably functioning individual memory system producing an inaccurate representation, or a representation appropriately causally connected to the past, that is shared but not influenced by interaction among misrememberers. This is just a proposal, however, and what types of memory errors these are and whether they are genuinely collective remains to be determined; nevertheless, I hope that my forays into this project serve as a catalyst for this kind of research.[22]

[22] I am grateful to audiences at the 2017 New Zealand Association of Philosophers' conference, the 2017 Philosophical Perspectives on Memory conference, the University of Otago philosophy postgraduate seminar, the 2018 Mental Time Travel: Origins and Function workshop, and the 2018 Australasian Association of Philosophy/New Zealand Association of Philosophers' joint conference for feedback that greatly contributed to the development of this chapter. I am also grateful to Zach Swindlehurst for proofreading.

CHAPTER 4:
MEMORIAL INJUSTICE

4.1. Introduction

The point of Chapter 2 was to develop a more inclusive account of testimony by proceeding from the observation that we can learn from others even when their testimony is non-propositional. The aim of doing so was to make sense of what was really at work when we learn from others. The present chapter approaches the same problem from a completely different angle: here, I continue to focus on how we might learn from others. Like the Mandela Effect advocates of Chapter 3, we are all more or less susceptible to incorporating information from others into our memories. While the Mandela Effect serves as a fun, bizarre example of this phenomenon, this chapter identifies an ethically and epistemically noxious way that we might learn from others. Perhaps most frighteningly of all, when it happens to us, we might not even realise that it is happening, and in some cases, the consequences can be grave.

Epistemologists who work on testimony are familiar with having to explain to laypersons that when we use the term "testimony," the phenomenon we describe is not restricted to the courtroom. Though legal testimony, of course, qualifies as testimony in the epistemological sense, there are many other kinds of communication that fall under the umbrella of testimony. Still, it is worth paying special attention to this legal sense of testimony, since what epistemologists say about it has stark practical ramifications for people's lives. In particular, testimonial injustice, when perpetrated against formal (i.e. courtroom) testifiers, can have grave practical consequences in addition to the epistemic consequences inherent to all forms of epistemic injustice.

The past decade or so has seen an explosion of literature on the topic of epistemic injustice, catalysed by Miranda Fricker's (2007) book *Epistemic Injustice:*

Power and the Ethics of Knowing.[1] Her conceptualisation of epistemic injustice as a "kind of wrong done specifically to someone in her capacity as a knower" and as taking the two distinct forms of testimonial injustice and hermeneutical injustice has been highly influential in the resulting discussion.

Fricker (2007) offers the following example of testimonial injustice: In the film *The Talented Mr. Ripley*, set in the 1950's, Marge Sherwood suspects that the disappearance of her fiancé Dickie is the responsibility of Tom Ripley. Upon expressing her suspicion to Dickie's father, Herbert Greenleaf, he dismisses her by saying, "Marge, there's female intuition, and then there are facts" (Minghella 1999). As it turns out, Marge was correct: Tom Ripley was indeed responsible for Dickie's murder, and Greenleaf's gender-motivated dismissal of Marge's concerns was a case of testimonial injustice. He explicitly invokes stereotypes about women as excessively intuitive and insufficiently rational in order to discredit Marge and ultimately reject her testimony. Had Greenleaf taken Marge as seriously as he would have a man, the police might have turned their attentions to Ripley sooner.

In testimonial injustice, a testifier lacks the social power required for their testimony to receive an appropriate level of credibility, and their testimony is therefore rejected by a more socially powerful recipient. The testimony is rejected on the grounds of prejudices related to the testifier's social type. Though I do not outline them here, Fricker offers other examples that are parallel to the case of Marge and Greenleaf, such as that of Tom Robinson's trial in *To Kill a Mockingbird*. Testimonial injustice brings about two kinds of harms: epistemic and practical. Primarily for Marge, for example, she is not taken seriously as a knower or giver of knowledge, explicitly because she is a woman. Secondarily, she suffers the practical harm of not receiving answers about her husband's disappearance. On Fricker's

[1] While Fricker's (2007) book sparked this explosion, many of the ideas contained therein were first articulated by other scholars—particularly women of colour. (See McKinnon (2016) for critical remarks on the genealogy of the concept of epistemic injustice.)

view, the epistemic harm is inherent in testimonial injustice, while the practical harm is not.

If what I claimed in Chapter 1 is correct and memory and testimony are indeed analogous, then the next step is to investigate whether analyses of downstream phenomena of one are applicable to the other. I began this project in Chapter 2 when I suggested that there was room to understand testimony-how as analogous to memory-how, and continued it in Chapter 3 by thinking about the interaction between memory and testimony in collectives. In Chapter 4, I focus on the phenomenon of *epistemic injustice* as it manifests in testimonial and memorial contexts. Epistemic injustice is a distinct kind of wrong that can be inflicted upon an agent specifically in her capacity as a knower (Fricker 2007). On Fricker's view, epistemic injustice comes in two forms: hermeneutical injustice and testimonial injustice. Though Fricker's work implies that hermeneutical and testimonial injustice exhaust the species of epistemic injustice, subsequent work has identified other distinct species, such as *contributory injustice* (Dotson 2012), *conceptual competence injustice* (Anderson 2017), and *interpretative injustice* (Peet 2017). In this chapter, I centre the discussion on testimonial injustice and bracket the others as promising areas for future consideration in light of what we know about the nature and epistemology of memory.

The aim of this chapter is to use the concept of testimonial injustice and the analogy between memory and testimony to develop an account of *memorial injustice*, in which—roughly speaking—a subject consciously or unconsciously rejects her own memories for unjust reasons. The exact details of this phenomenon will have to be worked out as the chapter progresses, but as a first approximation, let us consider an example of what memorial injustice might look like.

Suppose that a racialised suspect is arrested on suspicion of having committed a crime that he did not, in fact, commit. After extensive police interrogation, the suspect confesses, falsely, to the crime. The pressure and hostility

of the interrogation has engendered self-distrust such that the suspect discards his initial beliefs about the time in question and comes to believe what he is told by the interrogators. He comes to believe that he committed the crime even though he did not, and despite objective evidence that he is innocent (Chapman 2013).

Kassin and Wrightsman (1985) identify three kinds of false confessions: voluntary, coerced-compliant, and coerced-internalised. Voluntary false confessions are those in which subjects voluntarily incriminate themselves without external pressure, and can arise in the absence of an interrogation. These might be offered to gain notoriety or to shift suspicion off the guilty party. In contrast, coerced-compliant and coerced-internalised confessions both occur in interrogation settings. Coerced-compliant confessions happen when a subject is aware that the interrogators are convinced of his guilt, and so might plead guilty to avoid a long trial or to receive a reduced sentence, or simply to escape an unpleasant and aggressive interrogation. Finally, coerced-internalised confessions happen when "innocent people, subjected to misleading claims about the evidence, become confused, question their own innocence, infer their own guilt, and sometimes confabulate false memories to support that inference" (Kassin 2014, 114). Of the three types of confessions, the third is of special interest to us, for it requires a subject to come to adopt beliefs about his committing the crime. This is in contrast to the first two kinds, wherein the subject confesses but there is no corresponding doxastic shift: the suspect *knows* that he is confessing to a crime he did not commit. Indeed, as Gudjonsson and MacKeith (1982) claim, false confessions of this type occur when a subject develops "such a *profound distrust of their own memory* that they become vulnerable to influence from external sources" (Kassin et al. 2010, 15, emphasis added). In fact, a subject's questioning and ultimately rejecting his memories of the past and deferring to the accusations from the interrogators are a perfect example of memorial injustice.

For this example to be helpful, however, we shall need some clarification. Crucially, the individuals at stake in this phenomenon are not the interrogators and the suspect; rather, they are the suspect's past and present selves. That the suspect is also testifying to the interrogators is a red herring; the important phenomenon is whatever is happening in the suspect's internal mental life. In other words, what we care about is not the interaction between suspect and interrogator, but the increasing levels of internal epistemic self-distrust that manifest in the suspect's mind about his own memories. So, in the case of the false confessor, the memorial injustice is an experience *internal* to the suspect. The interrogators do not directly perpetrate the injustice, though their behaviour and questioning might catalyse it. In other words, if a suspect sincerely and accurately testifies that his activities on the night in question were such-and-such, and the interrogators disbelieve him, this is not an instance of memorial injustice, but of ordinary testimonial injustice. The memorial injustice obtains when the suspect himself begins to reject his *own* memories.

Importantly, memorial injustice is distinct from injustice with respect to testimony *about the past* – this would just be a species of testimonial injustice.[2] Because I focus only on individual memory, I bracket questions about collective memory (but see Tanesini (2018) for an account of collective memorial injustice)[3]. In this case, the rememberer will also have to be the perpetrator of the injustice, because others don't have access to her mnemic mental states except through testimony, and it will be too difficult to dissociate the epistemological problems with

[2] This line contains the implicit metaphysical assumption that individual memory occurs within one person. I comment on the metaphysics of a rememberer in Section 4. For now, however, as I have done in previous chapters, I bracket metaphysical concerns about characterising this view in terms of past and present self. I use it merely as shorthand to refer to the individual at the time of encoding and at the time of recall. I remain neutral on metaphysical concerns about personal identity.
[3]The term "memorial injustice" is also used by Tanesini (2018), but we have different (though compatible) meanings in mind. I use the term to refer to an internally-committed injustice, while Tanesini uses the term to refer to social rejection of collective memory, especially through the removal or destruction of artefacts (e.g., statues) that scaffold cultural or societal memory.

reports about the past that arise from their origin in memory from their expression in testimony.

In other words, if a rememberer reports on the past and a hearer rejects her testimony, it is not always obvious what causes the rejection: a hearer's identity prejudice (so an ordinary case of testimonial injustice), a hearer's rejection of memory as providing reliable grounds for knowledge about the past, or normal good epistemic reasons to reject testimony. For example, I might reject your testimony about the past not because I have assigned a credibility deficit to you, but because I am generally skeptical about the justificatory force of memory. In this case, it does not appear as though I have committed any injustice because your identity is not a factor in my assessment of your credibility as a testifier (in fact, your credibility has not come into at all; all that matters in this case is my belief that testimony is generally unreliable). Alternatively, I might fail to consider you to be a reliable rememberer; though I believe you are being *sincere*, I doubt that what you say is *true*. If this is the case, then what I am doing is doubting your reliability in testifying as to your memories; such a case is already adequately covered by Fricker (2007). Disentangling these phenomena is both difficult and unnecessary for my purposes, since the phenomenon at issue is not memory reports (i.e. testimony about one's memories), but an individual's internal mental experience of remembering and then accepting or rejecting her memories. In this chapter, for the sake of simplicity, I set aside the unique considerations about memory reports and epistemic injustice, since there are two stages at which an agent might judge whether she ought to accept a given memory or piece of testimony: one at the point of the rememberer's recall, and another at the point of the recipient's entertaining the testimony. The central case at issue for memorial injustice will therefore be an ordinary case of individual

remembering, where the source is the past self and the receiver is the present self.[4] In the case of our false confessor, it is important to note that the memorial injustice obtains not when the interrogators reject his testimony, but when he himself begins to doubt his own memories.

All this brings us now to the main point of this chapter: memorial injustice, I claim, is a species of epistemic injustice in which, *owing to a self-assigned identity-prejudicial credibility deficit or excess, a subject unfairly rejects or accepts a memory*. Importantly, the self-assigned identity-prejudicial credibility deficit or excess is an *internalised* assessment based on dominant views about members of the subject's social group. In the false confessor's case, owing to his internalised views about the (un)trustworthiness of members of his social group—further exacerbated by the hostile interrogation—he unjustly assigns himself a credibility deficit. He therefore distrusts his own memories, and ultimately comes to agree with what the interrogators already assume. This distrust is a direct result of internalising the negative stereotypes of members of his social groups, such that he ultimately distrusts and rejects his own memories, which he otherwise has a default epistemic entitlement to trust.

The thrust of the argument that I make in this chapter is this: memory and testimony are analogous, there is such a thing as testimonial injustice, so there is such a thing as memorial injustice, and memorial injustice comes about as a direct result of the memorial analogues of the phenomena that precipitate testimonial injustice. I proceed in this chapter as follows. In Section 4.2, I begin with a summary of the current state of the testimonial injustice literature, and show that no as yet identified species of epistemic injustice can accommodate memorial injustice. In Section 4.3, I recapitulate some of the claims I made in Chapter 1 about the

[4] Indeed, as interesting as it would be to look at the unique dimensions of memory reports and epistemic injustice, the project I have in mind here precedes such a discussion. Without understanding how an epistemic injustice can be fully internal, we cannot proceed to any meaningful discussion of how epistemic injustice plays out in the case of memory reports.

importance of mindreading to producing and interpreting testimony, and consider the ways in which metacognition in memory plays a similar role to mindreading in testimony, and argue that both are contributing factors to testimonial injustice. I then describe how memorial injustice might come about from specious metacognitive errors, which leads me to develop an account of memorial injustice toward the end of the section. Finally, in Section 4.4, I explore the possible harms that might result from memorial injustice.

4.2. Epistemic injustice

Fricker (2007) begins by defining epistemic injustice as occurring when someone is wronged in their capacity as a knower. She focusses on two kinds of epistemic injustice: *testimonial injustice* and *hermeneutical injustice*. The literature has developed considerably beyond the two forms of epistemic injustice identified by Fricker: notably, Dotson (2012) identifies *contributory injustice*, Anderson (2017) identifies *conceptual competence injustice*, and Peet (2017) identifies *interpretative injustice*.

I will elaborate on the details of each of these species of epistemic injustice shortly, but will begin by pointing out that all species of epistemic injustice have in common that they are instances of wronging someone *qua* knower. Indeed, this is epistemic injustice's defining feature. But beyond this defining feature, there are general features that all species of epistemic injustice share. Importantly, epistemic injustice is, at its heart, the result of the exercise of *identity power*.

In some cases, Fricker says, identity power, which leads to identity prejudices, is exercised by individuals (such as in the case of Marge Sherwood). But Fricker is careful to emphasise that if certain attitudes are pervasive and prevalent in a given society, there may be instances in which no agent need exercise power for a testimonial injustice to take root; that is, nobody needs to do or say anything *explicitly* to inflict the injustice on someone (15). If someone is silenced by the mere fact of her social type in such a way that she is prevented from testifying altogether,

she too suffers a testimonial injustice. Dotson (2011) calls this phenomenon *testimonial smothering*. The distinction between active and passive exercises of identity power shows that injustices can be agentially or socially inflicted, and that someone can suffer an injustice without anyone explicitly (even if silently) discounting her testimony. That injustices can be agentially or socially inflicted will also be significant in the upcoming discussion of memorial injustice, for it gives rise to the possibility of an agent falling victim to memorial injustice against herself, absent an external agent, by internalising identity power on behalf of—that is, in keeping with the interests of—a comparatively privileged group. In other words, as I will elaborate in Section 4.4, even if there is no external agent perpetrating the injustice against an agent (say, our false confessor), the prevalent structures of identity power might nevertheless bring about an epistemic injustice, including memorial injustice.

One feature worth noting is that each type of epistemic injustice epistemically harms not only the victims, but the perpetrators or the epistemic community as a whole, through harming individuals as knowers and routinely preventing the advancement of the epistemic community by limiting contributions from certain members of society. I identify such harms to the broader epistemic community as I proceed through the different species of epistemic injustice.

Now that I have specified the general features of epistemic injustice, the rest of this section is devoted to discussing the key features of each form of epistemic injustice: hermeneutical injustice, contributory injustice, interpretative injustice conceptual competence injustice, and testimonial injustice. As we shall see throughout Section 4.2, though memorial injustice is consistent with the general features of epistemic injustice, none of the as yet identified forms of epistemic injustice can adequately capture the uniquely memorial dimensions of memorial injustice.

4.2.1 Hermeneutical injustice

I address hermeneutical injustice first. For Fricker, hermeneutical injustice arises when, owing to an epistemic lacuna, members of a marginalised[5] group are prevented from contributing to epistemic resources. They are thereby rendered unable to comprehend—much less articulate—their own experiences of marginalisation.

Fricker (2007, 149–50) offers the example of Carmita Wood, who experienced sexual harassment in the workplace in the 1970s, before the conceptual resources to convey her experience had been developed. She left her job as a result of the sexual harassment, and, after being unable to find new employment, when she applied for unemployment benefits, she found herself unable to specify on the form why she had left her previous job. Women who had experienced sexual harassment had not been taken seriously as knowers and consequently been prevented from contributing to the pool of collective conceptual resources; thus, the language and concept to articulate the experience did not yet exist, so she could only write that the reason was "personal." Presumably, the unemployment office found the answer unsatisfactory, because Wood was subsequently denied unemployment benefits. Therefore, in addition to her epistemic marginalisation, she suffered the practical harm that was the financial blow of not receiving the unemployment benefit. Later, upon attending a consciousness-raising meeting, Wood learned that her experience in the workplace was not unique; collectively the group named the phenomenon "sexual harassment." Today, that critical concept is widely known.

Unlike testimonial injustice, which involves two agents (more on this later), hermeneutical injustice is a "purely structural notion" (Fricker 2007, 159), not

[5] For the sake of simplicity and consistency with the literature, I retain Fricker's language of marginalisation, but I wish to make it clear that what really matters is power. It is possible for an otherwise unmarginalized person to lack power in a certain social situation, and for that person not to become marginalized by that lack of power. For that reason, such a person might still be subject to an epistemic injustice absent marginalisation; in other words, marginalisation is not a necessary ingredient for epistemic injustice.

perpetrated by any one agent. It is a result of the cultural or societal failure of dominant groups to recognise marginalised persons as legitimate knowers, and to disregard their experiences such that the necessary conceptual resources to describe and articulate features of their existence, are absent. The injustice does not make itself known until a "more or less doomed attempt on the part of the subject to render an experience intelligible, either to herself or to an interlocutor" (Fricker 2007, 159).

Hermeneutical injustice has the peculiar consequence of epistemically depriving not only the marginalised group, but the dominant[6] group. In the case of Carmita Wood, for example, while she was able to articulate her experience to the other attendees of the consciousness-raising group, society at large remained ignorant owing to the hermeneutical lacuna where "sexual harassment" ought to have been. As a result, they failed to form (presumably true) beliefs about the world, although they did not encounter the same moral or practical harms that Wood herself did. Though the situation was likely worse for Wood, the hermeneutical injustice still epistemically disadvantaged those who could not understand her experience.

While hermeneutical injustice is a genuine form of epistemic injustice, it does not adequately capture what is going on in the false confessor's case. He does does not lack the appropriate conceptual resources to understand her own memories. The resources are in place and he has an adequate grasp on them, but he still fails to take his own memories seriously. So, since the reason he does not form the relevant beliefs is not a result of a lack of conceptual resources, the example is not a case of hermeneutical injustice. Thus, hermeneutical injustice cannot provide a way to understand the case. So, let us proceed to another species of epistemic injustice.

[6] Similar considerations to footnote 5 apply here: what matters is not dominance *per se* but comparative power. In this case, the dominant group is whatever group's is more powerful in a given situation, and is therefore better positioned to have their interests widely understood and recognised.

4.2.2 Contributory injustice

Where hermeneutical injustice arises when the culture lacks the appropriate conceptual resources to express the experiences of marginalised groups, *contributory injustice* (Dotson 2012) arises when the conceptual resources are available, and a marginalised person is able to comprehend and describe her experience, but the testimonial recipient's wilful ignorance prevents appropriate uptake. Like Dotson, Pohlhaus (2012) expands on Fricker's notion of hermeneutical injustice to advance an account of *wilful hermeneutical ignorance*, which falls within an "epistemology of ignorance" (Mills 1997, 18). Wilful hermeneutical ignorance refers to "instances where marginally situated knowers actively resist epistemic domination through interaction with other resistant knowers, while dominantly situated knowers nonetheless continue to misunderstand and misinterpret the world" (Pohlhaus 2012, 716). The link between wilful hermeneutical ignorance and hermeneutical injustice should be clear: widespread wilful hermeneutical ignorance amongst the dominant group will systematically produce hermeneutical lacunae. These lacunae, in turn, give rise to hermeneutical injustice whenever a marginalised person needs to invoke the very conceptual resources they have been prevented from contributing to the collective pool of knowledge.

Where contributory injustice differs from hermeneutical injustice is in the availability of conceptual resources. Hermeneutical injustice arises when a person tries to communicate something that relies on conceptual resources that (unjustly) do not exist; in contrast, contributory injustice arises when "an epistemic agent's willful hermeneutical ignorance in maintaining and utilizing structurally prejudiced hermeneutical resources thwarts a knower's ability to contribute to shared epistemic resources within a given epistemic community by compromising her epistemic agency" (Dotson 2012, 32). In other words, the necessary conceptual resources are available, but members of the dominant group fail to take seriously interlocutors who rely on those resources, and to engage with the resources themselves.

Let us revisit the case of the false confessor. We see here that contributory injustice will also not provide an adequate framework for understanding his situation. Recall that the reason the false confessor's case did not qualify as an instance of hermeneutical injustice was because the relevant conceptual resources existed, and hermeneutical injustices arise from epistemic lacunae. Hence, without a lacuna, there can be no hermeneutical injustice. This alone, of course, does not preclude his case from qualifying as an instance of contributory injustice, since contributory injustice involves wilful ignorance of the relevant resources, though the resources themselves are in place. For the internalised false confession to qualify as an instance of contributory injustice, the confessor would need to be trying to remember something, but her present (remembering) self would need to be wilfully misunderstanding the meaning of his memories. This does not seem to be the case, either, since it is not the case that he simply doesn't *understand* his memories, but that he does not *trust* them. It seems that he ought to be able to trust his own memories, and yet something in his social situation prevents him from what is otherwise a justified default epistemic self-trust (Burge 1993; Zagzebski 2012). In other words, absent a strong epistemic reason to reject his own memories, the best explanation for this rejection is that he has internalised a morally specious self-distrust. Since contributory injustice cannot account for the false confessor's case, we shall need to search further afield for an explanation.

4.2.3 Conceptual competence injustice

Anderson (2017) has identified another form of epistemic injustice: *conceptual competence injustice*. Conceptual competence injustice arises where "a member of a marginalised group is unjustly regarded as lacking conceptual or linguistic competence as a consequence of structural oppression" (Anderson 2017, 210). To illustrate the phenomenon, Anderson offers the example of a white, male, first-year graduate philosophy student who attends a philosophy of language seminar given by a philosophy graduate student who is a woman of colour. During the seminar,

the woman of colour asserts, "Natural kind terms are not rigid designators," which the first-year graduate student takes to be false. Based on his implicit beliefs about women of colour typically not excelling at metaphysics and the philosophy of language, and on his own understanding of Kripke's *Naming and Necessity*, the first-year graduate student assigns a lower degree of credibility to the presenter than he does to himself and judges the presenter to be less competent with the relevant concepts. As a result, the first-year graduate student fails to engage seriously with the speaker's arguments or presentation. Anderson puts the point nicely when he says that even though the presenter's understanding of the relevant concepts is more advanced, "the first-year student—under the influence of covert social norms of credibility ascription that propagate and sustain the epistemic oppression of women of colour and facilitate the epistemic privilege of white men – takes himself to have the greater conceptual competence" (Anderson 2017, 211). In other words, the first-year student endorses (even if subconsciously) the negative social expectations about the competence of women of colour when it comes to metaphysics and the philosophy of language, and thereby judges that he is more competent than she is.

Anderson points out that in doing so, the first-year student undermines himself epistemically because he fails to take on board an argument that ought to challenge his own understanding of the concept of rigid designation, as he would do if he were engaging with a presenter whom he judged to be at least as conceptually competent as he was. Moreover, he epistemically harms the speaker by treating her not "as a source of insight, but rather as someone who can be ignored" (211), and risks further damaging his willingness to assign an appropriate level of credibility to himself (he may overestimate his credibility) or to others (he may underestimate their credibility).

Here emerge some points of contact with the false confessor's case, although, as we shall see, not enough to convince us that conceptual competence injustice is the right way to understand it. The false confessor may well, for instance, have internalised and deployed negative implicit beliefs about his own trustworthiness,

such that he takes himself to be a less reliable knower than he really is. However, I doubt that conceptual competence injustice captures the whole story. For one thing, as the name suggests, conceptual competence injustice is primarily a matter of misjudging competence. In contrast, in the false confession case, the issue is not that the confessor judges himself to be incompetent with respect to the application of certain concepts. Instead, he judges himself to be untrustworthy because he is dishonest. Thus, conceptual competence injustice cannot explain the mental phenomena giving rise to false confessions, and we shall need to explore other species of epistemic injustice to find one that does.

4.2.4 Interpretative injustice

Where hermeneutical injustice concerns the existence of conceptual resources, contributory injustice concerns the availability of conceptual resources, and conceptual competence injustice concerns faulty assessments of competence with respect to the application of conceptual resources, *interpretative injustice* (Peet 2017) has nothing to do with conceptual resources whatsoever. Instead, "interpretative injustice occurs when the wrong content is assigned as a result of prejudicial stereotypes influencing interpretation, regardless of what concepts are available in our public language" (Peet 2017, 3427). In other words, interpretative injustice arises when prejudicial stereotypes cause a recipient to mishear or to misunderstand a testifier.

Peet claims that interpretative injustice is most likely where our interpretation is guided by stereotypes. This is especially so where the cognitive load on the recipient is the highest: for example, "in cases involving context sensitivity, loose talk, unfamiliar dialects or accents, noisy environments, and implicature" (Peet 2017, 3423). Assigning the wrong content to an utterance can take two different but related forms: first, a recipient may take the testifier to have *said* something other than she did (i.e. the recipient misunderstands or mishears the actual words of the testifier's

utterance), or second, a recipient may take the testifier to have *meant* something other than she did.

Here is a real-life example of interpretative injustice: in 2013, Ambridge Area High School in Pennsylvania went into lockdown when a receptionist misunderstood a student's outgoing voicemail greeting (Strauss 2013). The outgoing greeting was the student's own recording of the sitcom *The Fresh Prince of Bel-Air*'s theme song, which is a rap song that includes the lyric "Shooting some b-ball [basketball] outside the school." The receptionist who heard the outgoing message believed the student was saying "Shooting some people outside the school," and interpreted the message as a threat. The school went into lockdown, but when the misunderstanding was discovered, the lockdown was lifted and the student eventually cleared. Though reports indicate that the recording was somewhat garbled and unclear, the receptionist's predisposition to associate threats of violence with rap music[7] and to jump to the interpretation "Shooting some people outside the school" is consistent with Anderson's description of interpretative injustice.[8] (After all, if the message were very unclear, the receptionist would not have any special reason to jump to an interpretation of a threat of violence than any other, neutral interpretation.)

If[9] the student really did record "Shooting some b-ball outside the school," it is possible that owing to the association between rap music and violence, the receptionist heard "Shooting some people outside the school." In this case, the receptionist relied on prejudicial stereotypes about rap music and violence, which caused a misunderstanding of what the student's message really said. There were also practical consequences to this misunderstanding. The obvious consequence was

[7] There may also be a racial dimension to the assumption, since rap music is predominantly produced by Black artists in the US, and there are widespread stereotypes of Black men as violent or aggressive. However, as I was unable to find any information regarding the student's race, whether this association figured into this particular instance of interpretative injustice is unclear.

[8] In any case, even if the facts of this particular case are incorrect or incomplete, we can treat it as a hypothetical example illustrating interpretative injustice.

[9] I say "if" because there was never a definitive answer as to what the student had actually recorded.

that the school (and indeed, all schools in the county) went into lockdown. However, the student was also held accountable for a threat he did not make, and he was taken into police custody. The student's father pointed out that his son may have been cleared, but asked, "How is he supposed to go back to school and face his classmates?" (O'Shea 2013).

One obvious problematic consequence of interpretative injustice is that, like the student in the *Fresh Prince* debacle, victims of interpretative injustice are held accountable for assertions they did not make. Peet (2017) offers a more extended discussion of this possibility, but I will point out briefly that if a testifier claims that p, but is taken to have said q (where p is true and q is false), then the testifier may be unfairly judged to be incompetent in that domain. Moreover, if the same testifier routinely falls prey to this practice, then she may be judged to be incompetent across domains, or even globally incompetent. Alternatively (or additionally), she may be judged to be thoroughly dishonest or insincere. Neither incompetence nor dishonesty is desirable when one wants to be taken seriously as a knower.

The false confession, however, cannot be explained by interpretative injustice. This is because the subject's distrust in his memory does not arise from a misunderstanding of the *meaning* of the memory. There is no unjust or incorrect interpretation at play here. Rather, he interprets his memories correctly, but whatever he thinks he remembers, he rejects as wholly or at least substantially inaccurate. For the false confessor, the question is about the *content* of the memory itself, not its interpretation or what evidence it provides for his knowledge of the past; thus, interpretative injustice cannot fully explain his memory distrust or the consequent false confession. Since interpretative injustice fails, our best hope, I claim, lies in appealing to testimonial injustice.

4.2.5 Testimonial injustice

The point of the preceding subsections has been twofold: first, it has been to explore the state of the epistemic injustice literature. Second and more importantly, however,

the point has been to show that the example of what I call *memorial injustice* cannot adequately be described by the extant accounts of species of epistemic injustice. Given that—as I have argued in Chapter 1—memory and testimony are analogous, our best hope for understanding memorial injustice lies adjacent to *testimonial injustice*. Testimonial injustice has a few key features that require attention, since the remainder of the present discussion will be focused on testimonial injustice as it relates to the account of memorial injustice I develop toward the end of the chapter.

Importantly, testimonial injustice takes place between at least two different people. In a paradigmatic testimonial exchange, there is a testifier and a recipient. The participants in such an exchange will often occupy different strata of social privilege or power (for example, they may be of different genders, races, or classes). For Fricker, the root of testimonial injustice is a negative prejudice that the hearer holds about some aspect of the speaker's identity that causes the hearer not to grant the speaker's testimony the credibility that it would otherwise merit. In Fricker's language, this is known as an "identity-prejudicial credibility deficit" (2007, 28).

Most fundamentally, testimonial injustice occurs when a speaker's credibility is diminished due to a prejudice or prejudices held by the hearer, or in other words, when a speaker's testimony is rejected on non-epistemic grounds. In cases such as these, were the same testimony to come from another speaker who is not a member of a marginalised group, it would not be rejected unless there was something in the *content* of the testimony itself to render it suspect. In Fricker's words, testimonial injustice is an "*identity-prejudicial credibility deficit*" (28, emphasis added), meaning that due to a prejudice pertaining to the identity of a testifier—for example, their race, gender, or class—a hearer wrongly assigns the speaker a credibility deficit. That is, a hearer takes a testifier to be less credible than she really is simply because of the testifier's social type. In having such a credibility deficit assigned to her, the force of the testifier's testimony is damaged and the testimony is consequently rejected. In other words, the testimony is rejected not for good epistemic reasons

pertaining to the content of the testimony, but instead for non-epistemic reasons pertaining to the identity of the testifier.

To say that the central case of testimonial injustice is a matter of credibility deficit is not to say that the *only* cases of testimonial injustice are matters of credibility deficit. Identity prejudices can also cause credibility excesses, where a testifier is granted more credibility than she ought to be. A senior academic who asks a junior colleague for feedback on a paper draft may unfairly be assigned a credibility excess by the junior, who then does not offer comments as critical or incisive as they would have been if he had engaged with someone he took to be a true epistemic peer. In this case, the senior academic suffers epistemically due to an unfairly inflated level of credibility (Fricker 2007, 18–19). Fricker, however, interprets this as a one-off error that is unlikely to have broader epistemic or social consequences for the senior academic; indeed, more often than not, credibility excesses are more advantageous than they are disadvantageous. While credibility excesses may place an undue or undesired epistemic burden on the recipient, credibility excesses do not constitute the same epistemic insult or devaluation of epistemic personhood that credibility deficits do (Wright forthcoming). Fricker then sets aside credibility excess and maintains that *identity-prejudicial credibility deficit* is at the heart of the central case of testimonial injustice.

Testimonial injustice offers us the best starting point for understanding memorial injustice, but it alone cannot explain phenomena like gaslighting or false confessions. Typical cases of testimonial injustice do not change the testifier's mental states; while the recipient fails to give appropriate uptake to the testimony, in a typical case, this failed uptake does not affect the testifier's credence in the proffered testimony. Gaslighting and false confessions, however, *do* change the mental states (or at minimum, their doxastic attitudes) of the testifier by bringing about epistemic self-distrust, and testimonial injustice alone cannot explain how this is the case. So, while memorial injustice may have strong links to testimonial injustice, testimonial

injustice still fails to offer a satisfactory account of the phenomenon. Yet, it seems clear that an epistemic injustice of some kind has taken place, since the agents are wronged distinctly in their capacities as knowers. When an aggressive interrogation, for example, brings about epistemic self-distrust such that a suspect comes to believe (falsely) that he committed the crime in question, there is certainly some harm to the suspect's status as an epistemic agent. Not only does he suffer a testimonial injustice from the interrogators, he suffers a damaging blow to his assessment of his own credibility—and one that has potentially catastrophic consequences, such as wrongful conviction.

The rest of this chapter is devoted to developing an account of how memorial injustice arises and the implications it has for epistemic agents, and epistemology writ large. The account will also be able to characterise phenomena such as gaslighting and internalised false confessions. Before I progress to developing this account, however, it is worth dedicating some space to discussing credibility excesses and deficits in greater detail, for they will play a significant role in the following argument.

4.2.6 Credibility excess and deficit

The lack of attention Fricker pays to credibility excesses has not gone unquestioned in the subsequent literature. José Medina (2013) takes issue with Fricker's (2007) characterisation of the central case of testimonial injustice as being a matter of credibility deficit, and along with Elizabeth Anderson (2012) points to credibility *excess* as posing just as serious a threat to epistemic practices as do credibility deficits. "Prejudice," Medvecky (2018, 1397) says, "can also be positive prejudice—it can be unreasonably supportive of some individual's knowledge." In other words, it might artificially lend credibility to a testifier who would not otherwise deserve it. For example, when high-profile scientists Neil deGrasse Tyson and Bill Nye made disparaging remarks about the value of philosophy, those who took them seriously overestimated Tyson's and Nye's knowledge about philosophy. It is true that Tyson

and Nye are experts in science, but neither has any special knowledge about philosophy and so are not in a position to make authoritative claims about it. Nevertheless, their status as scientists gave rise to positive prejudices that caused inflated credibility assessments; thus, they were more likely to be believed when they made claims about philosophy. Genuine credibility in one domain (science) granted them artificial credibility in another (philosophy). This is how credibility excesses work.[10]

Credibility excesses can be damaging to epistemic practices in several ways: first, like the senior academic who does not get sufficiently critical comments from the junior colleague, credibility excesses can hamper an individual's epistemic progress. Second, credibility excesses allow privileged testifiers to make claims with lower levels of scrutiny than they would otherwise require, thereby increasing the risk of spreading misinformation. Third, they quash the claims made by less-privileged testifiers who offer testimony that contradicts, competes with, or is otherwise incompatible with the testimony of testifiers who are assigned credibility excesses. In this third sense, credibility excesses can thus lead to the unjust rejection of less-privileged testifiers' testimony, even if the recipient holds no negative prejudices about the social identity of a less-privileged testifier.

Both credibility excesses and deficits come about as results of prejudices. Unsurprisingly, credibility deficits often result from negative identity prejudices held by the testimonial recipient, while excesses result from positive prejudices. The next question, then, is where these identity prejudices originate. One prominent suggestion has been that stereotypes play an important role in the development of identity prejudices. Let us consider, alongside Fricker, how something like stereotyping might affect a testimonial exchange. Stereotypes are, generally

[10] The phenomenon of experts in one field wading into another field in which they are not an expert and passing judgment nevertheless has been called "epistemic trespassing" (Ballantyne 2019) and "cross-expertise poaching" (Nichols 2017). I think the role of credibility excess in this phenomenon is both apparent and fascinating, but I leave the question for future research.

speaking, widely-held collective assumptions about the natures of certain groups, and are typically deeply resistant to counter-evidence (Blum 2004). Importantly, though, an agent need not endorse, agree with, or consciously accept a stereotype for them to fall prey to its normative force, and this is true whether or not the agent is a member of the group targeted by the stereotype.[11] For example, a woman is not immune to internalising sexist stereotypes simply because she is a woman, nor is she immune to deploying them against herself or other women (Bearman, Korobov, and Thorne 2009; Liebow 2016; Szymanski et al. 2009). What this means is that, if stereotypes are at the root of identity-prejudicial credibility excesses or deficits, then a member of a targeted group is not thereby insusceptible to making an unjust credibility assessment of a fellow member of that group. This fact will be important later, when we begin to make sense of how it could be the case that an individual could be unfairly biased against (or for) herself simply in virtue of her membership in a targeted group.

The preceding discussion of credibility excesses and deficits gives rise to an objection. You might think that to say there are such things as credibility deficits or credibility excesses requires positing that there is such a thing as a correct or appropriate degree of credibility. You may also think that determining whether an agent has suffered a credibility deficit or been granted a credibility excess cannot be done if the appropriate degree of credibility is unknown. There are three points to be made in response. First, this objection runs the risk of conflating the epistemological with the ontological, when they ought to be kept separate. The question of *whether* an injustice has occurred is different—but related—to the question of what the appropriate method is for *determining* whether an injustice has occurred. In other words, even if we have no way of identifying the appropriate level of credibility to assign to a testifier, this does not entail that there *is* no appropriate level of

[11] For more in-depth discussions of this phenomenon, see works on stereotype threat and implicit bias, e.g., Gendler (2011), Saul (2013), Steele (1997), Steele and Aronson (1995; 2004), and Steele, Spencer, and Aronson (2002).

credibility; to conflate the two is to conflate the epistemological question with the ontological question.

Second, the notions of credibility and excess here are relative. Where Fricker (2007) argues that credibility is not a *distributive* good, Medina (2011; 2013) responds that credibility is a *comparative* good. To see the difference, consider this analogy: height is not a distributive good, in that for one person to be tall does not require that another person be short; nevertheless, the concept of height can still serve as a point of comparison between the two. Credibility, Medina claims, is the same. Just as there need not be any standard or norm of height in order for us to understand that some people are tall and some people are short, the same is true of credibility: in many cases, we will be able to tell whether someone has enjoyed a credibility excess or suffered a deficit even in the absence of a reference to an independent standard.

Third, even if we *were* to say that there were some standard or appropriate level of credibility, the problem of vagueness arises here. Here, again, an example will be helpful. Let us compare credibility levels with voting: clearly, a toddler is too young to vote. On the other hand, surely a 35-year-old is old enough to vote. Knowing that the toddler is too young and the 35-year-old old enough does not necessarily entail knowing exactly what the "right" voting age is.[12] Similarly, in the credibility case, neither identifying clear cases of excess and deficit nor making claims about them requires knowing or being able to determine exactly what the appropriate level of credibility is: the issue a matter of degrees. The issue of whether there really is such a thing as an appropriate or correct level of credibility is an issue that certainly requires an extended philosophical treatment and analysis, but for now, I follow other epistemic injustice scholars in bracketing the question. Any work on epistemic injustice that relies on the notions of credibility excess or deficit will

[12] I owe this example to Fabien Medvecky.

eventually have to answer to the question of appropriate levels of credibility, but doing so is not my project in this chapter.

I leave the questions of whether there is such a thing as an appropriate level of credibility, and how to determine what it is, as directions for future research. I hope it is enough to say, at this point, that a credibility excess arises when an agent is granted more credibility than she would have been given if she were another speaker of a different but epistemically irrelevant[13] social type, and a deficit when she is granted less, and that these misattributions of credibility are the result of prejudices on the part of the receiver toward an aspect of the source's identity. From this point to the end of the chapter, I take for granted that there is such a thing as a testimonial injustice, and that it is a matter of having one's testimony unduly rejected owing to an identity prejudice on the part of the hearer.

The point of this long foray into the species of epistemic injustice may seem tangential or irrelevant, but I doubt this is genuinely the case. Importantly, no species of epistemic injustice that has been identified so far in the literature can completely and adequately capture and characterise the unique phenomena at issue in memorial injustice. Though some of the issues at stake are not entirely unique to memorial injustice, the phenomenon stands alone in that it is internal to one epistemic agent, although it is true that the actions of external agents, such as hostile interrogators, can precipitate or exacerbate the effects of memorial injustice. The similarities between memorial injustice and other species of epistemic injustice suggest that memorial injustice can move into the house, so to speak; the differences suggest it should get its own room.

There is, however, one final feature of memorial injustice that must be highlighted before we can proceed. Importantly, one thing that distinguishes

[13] For considerations on whether and when a knower's social types are epistemically relevant, see the literature on standpoint epistemologies, e.g., Code (1981), Collins (1986; 1990), Harding (1991; 1992), Hartsock (1983), hooks (1984) and Smith (1974; 1987).

memorial injustice from other species of epistemic injustice is that its epistemic harms are even more acutely targeted: while epistemic injustice, broadly speaking, is always a matter of wronging someone in her capacity as a knower, memorial injustice is unique in its potential to precipitate a doxastic shift in victims. In other words, memorial injustice has the power to alter the beliefs of the victim. Consider the innocent false confessor who, despite evidence to the contrary, and despite having no recollection of committing any crime, incriminates himself after a long and hostile interrogation. What makes this case different to, say, a garden variety testimonial injustice is that in the latter case, a subject might testify and have her testimony rejected, but the rejection does not alter her doxastic attitudes. Memorial injustice is different. It arises when a subject comes to doubt her own memories so profoundly that she rejects them, and may, instead, rely on external cues to fill the void that remains. In false confession cases, for example, subjects who suffer from memory distrust (Gudjonsson 2003) are more susceptible to accepting misinformation and confabulating harmonious representations than their counterparts who are optimistic about the reliability of their memories (van Bergen et al. 2010). This phenomenon exemplifies the kind of doxastic shift I mentioned above: the involuntary modification of one's own memories to accord with cues from external sources.

The upcoming discussion proceeds from the claim that, as I argued in Chapter 1, metacognition and mindreading play analogous roles in remembering and testifying, respectively. I proceed from Hyde's (2016) claim that testimonial injustice is at its core a failure of mindreading, and apply the claim, *mutatis mutandis*, to metacognition in memory. Ultimately, I show that metacognition can fail in ways analogous to those in mindreading, and that those failures can be both epistemically and morally specious. The basic thrust of the argument is that if an agent rejects her own memories on the basis of a morally or epistemically specious failure of metacognition (in a way analogous to a failure of mindreading in testimony), then

she has perpetrated a memorial injustice. Starting in the next section, I develop an account of memorial injustice, beginning with the relationship between metacognition and mindreading.

4.3. The case for the existence of memorial injustice

In Chapter 1 of this thesis, I argued that memory and testimony are analogous. Principal support for this claim comes from the evidence that both memory and testimony are reconstructive processes, where cognitive labour is performed by source and receiver alike to produce a representation or message fit for uptake. I also argued in Section 1.4 that a variety of noise-like phenomena can contribute to what is remembered; in this chapter, mindreading and metacognitive credibility assessments might fill this niche. The purpose of the present section is to revisit the analogy between memory and testimony to provide support for the argument in what remains of this chapter. Specifically, I wish to revisit the question of mindreading and metacognition and the role each one plays in producing a final message. To this end, I consider the types of errors one might make during the production or interpretation of messages, especially errors pertaining to mindreading and metacognition. The aim of this section, then, is to begin to work toward developing an account of memorial injustice.

The first step is to explore what kinds of error lead to credibility deficits and excesses. Mindreading is the obvious target, since it is through mindreading that we assess others' background knowledge, motives, desires, intentions, and, importantly, credibility. Thus, errors in mindreading can produce erroneous credibility assessments, thereby precipitating testimonial injustice. I proceed in this section as follows: first, I revisit those features of the memory-testimony analogy that encourage us to turn our attention toward mindreading and metacognition. Second, I describe the relationship between metacognition and mindreading, and argue that we are justified in thinking that the two are relevantly similar when it comes to

epistemic injustice. Third, I consider evidence about error in metacognition and mindreading. Finally, I advance an account of memorial injustice.

Let us begin by returning to the memory-testimony analogy. In memory, the past self at the point of encoding is responsible for selection, abstraction, interpretation, and integration, while the present self is responsible for reconstructing memory content (Alba and Hasher 1983). In testimony, the testifier is responsible for performing an act of communication, and the recipient is responsible for interpreting that act of communication (Grice 1975; Lackey 2008; Scott-Phillips 2015; Sperber and Wilson 1986). In both cases, source and receiver make judgments about the other's background information, mental states, motives, and competencies that influence whether and how the memory or testimony is taken up. Making these judgments requires a great deal of cognitive labour, both conscious and unconscious. Crucially, mindreading in testimony and metacognition in memory contribute important information that influences the resultant belief (Michaelian 2011c; 2012a; Scott-Phillips 2015; Shanton and Goldman 2010).

When it comes to memory and testimony, one need not accept that mindreading and metacognition just *are* the same cognitive process, or even that they are different applications of the same cognitive mechanism. One need only accept that mindreading and metacognition play relevantly similar roles in forming beliefs; for that reason, the relationship between the relevant cognitive mechanisms (or the relationship between the applications of the same mechanism) is somewhat inconsequential. Nevertheless, it is worth spelling out a specific viewpoint on the issue.

One view is that third-person mindreading is prior to first-person metacognition; in other words, metacognition is simply the result of turning our mindreading capacities back on our own minds (Carruthers 2006; 2009; Gazzaniga 2009; Gopnik 1993; Wegner 2002; Wilson 2002). As Carruthers (2009) puts it, this view entails that there is no significant difference between metacognition and

mindreading, since the two are simply the same capacity with different targets. If this is right, then we should observe similar failures in both mindreading and metacognition. This view emerges against a background of ongoing intellectual debate rather than consensus, however, and it is important to acknowledge that if mindreading and metacognition are wholly different mechanisms, we have much less reason to think that the two should behave analogously. Nevertheless, the view that they are entirely different is a minority view, and there is general convergence on the point that either mindreading and metacognition are two different applications of the same mechanism, or that one underpins the other. I encourage the concerned reader to consult Carruthers' (2009) comprehensive review of the different views of the relationship between metacognition and mindreading to assuage those worries. The next two subsections are dedicated to exploring how error in mindreading and metacognition figure into epistemic injustice.

4.3.1 Mindreading error in testimony

Coming to accept another's testimony requires several components to fall into place. First, the testifier must perform some act of communication; second, the act of communication must be intelligible to the recipient; and third, the recipient must judge the content of the testimony to be worthy of acceptance. This last step requires, in part, that the recipient judge the testifier to be sufficiently credible. As Hyde (2016) puts it, "since testimonial injustice involves a failure to accurately perceive another as credible, we ought to consider such failures as failures of mindreading." She also cites evidence that mindreading is crucial to social interaction (Gallese 2001; Goldman 2005; Meltzoff 2005). Hyde claims that the faulty credibility assessments that engender testimonial injustice result from breakdowns of mindreading.

One initial worry here is that appealing to a failure of mindreading might only explain *some* cases of testimonial injustice. Ostensibly, I can read on your face whether you are being dishonest (though I might be incorrect in my conclusion), but there does not seem to be a good reason to think that I can similarly read on your

face that you are competent in the relevant domain. So, something else must be able to explain the remaining cases. If this is correct, then the argumentative move I am trying to make here—to say that memorial injustice is akin to testimonial injustice—is in jeopardy, because the failures of metacognition and mindreading are disanalogous.[14] If mindreading failure can only explain testimonial injustice arising from erroneous assessments of another's *sincerity*, then we have no principled reason to think that analogous metacognitive failures will do anything to explain memorial injustice based on erroneous assessments of one's own *competence*.

Responding to this objection will require a little more understanding of what mindreading is and how it works. Mindreading is the cognitive capacity (or set of cognitive capacities) by which we ascertain others' mental states, intentions, beliefs, desires, and so on (Ravenscroft 2016). On one understanding, mindreading is deeply vision-based. Successful mindreading is done by observing facial expressions and gestures, and interpreting those in a way that supports an inference about the speaker's mental states. For example, if I see a person injure themself and cringe in pain, I too might cringe in response. This is an example of low-level mindreading (Hyde 2016; Shanton and Goldman 2010). However, there also exists high-level mindreading, which is "more complex [than low-level mindreading] and tends to involve propositional attitudes" (Shanton and Goldman 2010, 531). In simulation-based high-level mindreading,[15] subjects isolate or temporarily discard their own beliefs, generate a series of pretend inputs about another person's mind, run a simulation, and produce a pretend output. This kind of mindreading does not necessarily depend on visual cues in the way that low-level mindreading does. Where, for me to have the right sort of mirror neuron reaction—the kind integral to low-level mindreading—when I see you drop a hammer on your foot requires me to

[14] Thanks to Jordi Fernández for raising this worry.

[15] The difference between low- and high-level mindreading is orthogonal to the contest between the *theory theory* of mindreading (Churchland 1988; Fodor 1987; Sellars 1955) and the *simulation theory* of mindreading (Goldman 1992; Gordon 1986; Harris 1989; Heal 1986). Both theory theory and simulation theory agree that low- and high-level mindreading exist.

see you cringe and then generate my own response to your pain, high-level mindreading depends instead on my information about your initial states, such as your beliefs and desires. I can ascertain these states in myriad ways, and thus I do not necessarily depend on witnessing your facial expressions or gestures to obtain a set of inputs for my mental simulation. I can attribute certain mental states to you that afford you more or less credibility, perhaps because I attribute to you more or less competence in the relevant domain. Thus, we can avoid the above worry about whether appealing to the role of failures of mindreading in testimonial injustice will provide a suitable argumentative route to the role of failures of metacognition in memorial injustice.

Note that high-level simulation-based mindreading can produce incorrect outputs "for a variety of reasons. A mind reader might lack pertinent information about his target's initial states (preferences, beliefs, and so on) or he might fail to 'quarantine' or inhibit his own genuine states when doing a simulation" (Shanton and Goldman 2010). These factors might lead a person to attribute lower credibility to a testifier than she should, since she might fail to take her seriously as a competent knower.

Hyde claims that even low-level mindreading is not immune to error, pointing out that while newborn babies show no preference for race, by three months of age they exhibit a preference for own-race faces, and by nine months of age some infants' facial recognition works only within their own race. She claims, following Kelly et al. (2005) and Bar-Haim et al. (2006), that this narrowing of facial recognition ability is acquired and results from limited exposure to and interaction with members of other races. Hyde says, "if newborns maintained [their initial] lack of race-preference as their mindreading abilities develop, then it is possible that they may be less affected by social stereotypes, both because of their own mindreading abilities, and because of their having regularly experienced persons of many races fulfilling other-than-stereotypical behaviour" (865). By adulthood, people are in fact

less capable of low-level mindreading of members of other races, which, Hyde claims, contributes to testimonial injustice.

Now that we have examined how failures of mindreading can provide the appropriate conditions for testimonial injustice to take root, let us consider how similar failures of metacognition can give rise to memorial injustice, including internalised false confessions.

4.3.2 Metacognitive error in memory

Recall from earlier in this chapter that I hold that mindreading and metacognition play sufficiently similar roles in testimony and memory, respectively. Importantly, metacognition plays two roles: first, it regulates cognition, and second, it provides knowledge of the contents of cognition (Moshman 2018; Nelson and Narens 1994; Schraw and Dennison 1994; Schraw and Moshman 1995). Metacognition can contribute to whether an agent endorses or rejects the outputs of cognitive processes, including memory (Michaelian 2012b); this will become particularly significant when we consider the role of metacognitive error in producing memorial injustice.

The upshot for the present section is that if testimonial injustice is a failure of mindreading, and metacognition is to memory as mindreading is to testimony, then memorial injustice is a failure of metacognition. Otherwise put, memorial injustice arises when a subject makes a metacognitive error that results in the unjust rejection (or endorsement) of memory outputs.

Metacognition, like perception, is an innate capacity, but is not thereby immune to influences from culture or theory. Proust and Fortier (2018) identify metacognitive variations across cultures, suggesting that while metacognition is a natural cognitive capacity, the roles metacognition plays, the epistemic norms it enforces, and the errors for which it monitors vary greatly across (and within) cultures. Cross-cultural variation aside, the outputs of metacognition vary greatly depending on a variety of other factors. Interestingly for our purposes,

metacognition varies on affective state (Efklides 2015): students who were in a negative mood were less confident that they had achieved their goal score on an exam (Lane et al. 2005). In an experimental setting, van Bergen (2011) found that negative feedback from an experimenter increased students' memory distrust and internalised false confessions. What this means for the false confessor is that in guilt-presumptive interrogations where suspects endure sustained hostility, it may be only a matter of time until memory distrust creeps in and false confessions become more likely.

Perhaps the best way to think about the role of metacognitive failures in memorial injustice is to adopt the source monitoring framework analysis of false confessions, proposed by Henkel and Coffman (2004). Proceeding from observations from Kopelman (1999) and Schacter (1999), Henkel and Coffman claim that internalised false confessions may be due, at least in part, to source monitoring errors. Source monitoring is the ability to identify the origin of a particular mental representation (Johnson, Hashtroudi, and Lindsay 1993). For instance, I can usually tell whether I am remembering something or imagining it.[16] I can also remember who told me that joke or in what tabloid I read that gossip column, whether I know so-and-so's philosophical argument because I read it in a journal or saw them give a presentation, and so on. All of these capabilities are applications of source monitoring. Henkel and Coffman argue that internalised false confessions come about because suspects incorporate the testimony of the interrogators into their representations of the past, and misattribute the information to their memory. The phenomenon is not limited to internalised false confessions, either: many of the therapeutic techniques employed in the recovery of so-called "repressed memories" of childhood sexual abuse are now known to increase the likelihood of false memory generation (Henkel and Coffman 2004; Loftus 1997a; 1997b).

[16] Maybe. It's complicated (De Brigard 2014; Michaelian 2016a).

To compound matters, according to Gallo and Lampinen (2015), retrieval monitoring itself—the "search and decision process that people use to regulate accuracy at the time of retrieval" (389)—is influenced by one's beliefs about one's own metamemory capacities (where metamemory is metacognition of memory). One suggestion that Gallo and Lampinen have for preventing false memories is that remembering subjects seek corroboration of their memories; however, this will prove difficult in an interrogation context, where hostile interrogators not only fail to corroborate a suspect's memories, but present conflicting information instead. Moreover, what this means is that if one has already had one's ability to remember accurately challenged, one might already have suffered a blow to his beliefs about his own metamemory capacities. If this is the case, then subjects are doubly damned: not only are they making a source monitoring error, they are more likely to make a source monitoring error because their confidence in their source monitoring abilities is under threat.

Shaw and Porter (2015) were able to generate, in a controlled laboratory setting, full episodic false memories of committing a crime. In some cases, participants accept that they are guilty due to the new, introduced "evidence," though they still have no memory of the crime. In more extreme cases, the introduced information provides the basis for confabulated details of the crime, thereby generating a false memory (Shaw and Porter 2015). There are some obvious connections here to my claims about collective confabulation in Chapter 3 of this thesis, and about the integration of information from other sources into one's memory; indeed, as Sigurdsson and Gudjonsson (1996) found, subjects who were more prone to confabulation were also more prone to internalised false confessions. Compared to the Mandela Effect, however, here the mental intrusion seems to be much more sinister. (I return to this point in Section 4.4 of this chapter.)

So far in this section, I have shown that metacognition can fail in a way that allows confabulated incriminating representations to stand in for genuine

exculpatory memories. So, it would seem that while metacognitive error is not solely responsible for bringing about memorial injustice, it is at least a key component in it.

4.3.3 An account of memorial injustice

The aim of this subsection is to make clear exactly what I take memorial injustice to be. So far, I have said that memorial injustice is the result of a failure of metacognition. But just as the presence of mindreading error alone does not a testimonial injustice make, neither does metacognitive error guarantee memorial injustice. However, where self-directed credibility assessments are deflated by internalised prejudices about groups of which the subject is a member, and by the ethically toxic behaviour of an authority figure, the likelihood that an injustice has arisen is higher. In other words, we might think of morally specious metacognitive failure as a necessary but not sufficient condition for memorial injustice.

An important clarification here is this: neither Fricker (2007) nor I mean to assert that *every* time an agent who is a member of a group targeted by negative identity prejudices has her testimony rejected, she has automatically suffered a testimonial injustice. There are, of course, many legitimate reasons to reject testimony, and these reasons can be deployed against all kinds of testifiers, regardless of their membership in a group targeted by identity prejudices. What distinguishes an ordinary, non-culpable rejection of testimony from a testimonial injustice is whether the hearer's prejudices have played an epistemically culpable role. This can be spelled out counterfactually, if crudely: if a receiver rejects the testimony of an agent who is a member of a group targeted by identity prejudices, when she would have accepted the testimony had the testifier not been a member of that group, the receiver has perpetrated an epistemic injustice. This counterfactual formulation, however, will fail to capture every instance, for there are also instances in which one suffers a credibility deficit but it does not put her testimony below the threshold for acceptance. Her credibility may be damaged but her testimony not

thereby rejected. Thus, the counterfactual formulation should be taken as a sufficient but not necessary condition for testimonial injustice.

There are multiple reasons that a speaker might be assigned diminished credibility, but not all of them constitute an injustice. For instance, if a credibility deficit is traceable to non-culpable[17] misunderstanding or error, consequent rejections of testimony are less likely to constitute an injustice (Wright forthcoming). Determining which instances of assignments of credibility deficits rise to the level of injustice depends crucially on whether the assignment is just such a misunderstanding or error, or whether it is a result of an unjust identity prejudice. But without perfect access to the mental states of the person who assigns a credibility deficit, determining whether it is an injustice or merely an error is difficult. An inability to determine this, however, does not change the fact of the matter.

For the present section, let us set aside the question of mere (i.e. non-culpable) error and explore cases of error arising from unjust identity prejudice. If we are to say that some cases are merely erroneous and some are unjust, then the question of which cases are which matters for determining which are genuine cases of memorial injustice. In making this distinction, the first thing to note is that all injustices involve error, whether moral or epistemic. However, in judging the blameworthiness of the agents involved in the injustice, it is important to determine whether there is such a thing as a non-culpable error. The answer to this question essentially comes down to whether people are responsible for their biases (Holroyd 2012; Holroyd, Scaife, and Stafford 2017). If people are indeed responsible for their biases, then their actions in the injustice will inherit the culpability. If, however, people are not always

[17] The non-culpability requirement here is important, for it keeps perpetrators of such wrongs as interpretative injustice or conceptual competence injustice on the epistemic (and moral) hook. Wilful or otherwise morally specious misunderstanding remains squarely in the domain of epistemic injustice, especially if such misunderstanding is due to identity power or identity prejudice.

responsible for their biases, then errors arising from biases might not always be culpable.

What this means for the metacognition argument is this: one idea central to the notion of responsibility is that we are not responsible for things outside of our control. Thus, you might think that people haven't committed an injustice if they've made some kind of mistake with their metacognition, because metacognition is usually beyond our conscious control.[18]

Nevertheless, even if the suspects are not responsible for their own biases, and so not morally culpable in the false confession, the suspect's biases are not the only morally specious element of memorial injustice. We must also attend to the fact that hostile, guilt-presumptive interrogators are providing the necessary background conditions for memorial injustice to obtain. So, in short, memorial injustice arises when morally specious background conditions provide fertile ground for memory distrust to take root, and when that memory distrust is so profound that it causes the outright rejection of one's own memories and, contingently, the internalisation of their morally noxious alternatives.

4.4. Harms and implications of memorial injustice

Above, I developed an account of what memorial injustice is: namely, an instance in which, owing to a self-directed prejudice, one errs in assigning credibility to oneself such that he overestimates or underestimates the epistemic worth of his own memory. The more insidious forms of memorial injustice will involve credibility deficit, since it is likely that one's epistemic (and psychological) confidence will suffer as a result, and may further corrode his willingness or ability to contribute to future epistemic projects. This is not to say, however, that credibility excesses do not

[18] This point aligns with Hyde's criticism of Fricker's "testimonial sensibility": Where Fricker recommends developing a "testimonial sensibility" to counteract identity prejudice and minimise occurrences of testimonial injustice, Hyde suggests that proposing such a testimonial sensibility is overly optimistic with respect to agents' ability to control their metacognition (Hyde 2016; see also Sherman 2016).

also pose an epistemic threat. The Dunning-Kruger Effect (Dunning 2011), for example, arises when individuals are unaware of their own ignorance and attribute a credibility excess to themselves, such that they overestimate their own competence and thereby deflate their assessments of the credibility of those around them.[19] This section is devoted to exploring these kinds of implications of memorial injustice, but before I begin to discuss them, there is an obvious objection about the metaphysics of remembering: that it is not clear whether the present self and past selves are really so different. First, I address this objection. Second, I explore the practical and epistemic harms that might result from memorial injustice. Finally, I consider how memorial injustice might shed light on existing phenomena.

4.4.1 Concerns about past self and present self

At this point, the objection may be raised that, while testimonial injustice is (at least on Fricker's view) a two-agent phenomenon involving both testifier and recipient, remembering takes place internal to one subject. If this is true, then we arrive at a puzzle for epistemic injustice. If there is only one subject, then it is much harder to see how an agent could perpetrate an injustice against himself, but if there are two agents (past self and present self), then it is hard to see where the harm comes in; surely the past self is impossible to harm, since he is no longer accessible. Let us explore this puzzle in more detail below.

By way of first response, it seems as though epistemic injustice is not necessarily agential at all. Hermeneutical injustice, for example, is a purely structural—rather than agential—notion (Fricker 2007, 159), where no one individual perpetrates the injustice, although individuals and collectives alike are harmed. Instead, hermeneutical injustice is produced by features of society that enable participation in the epistemic life of groups for some members of society more than others. So, it is not necessary for epistemic injustice that there be an agent:

[19] See also Wright (forthcoming) on epistemic harm and the Dunning-Kruger Effect.

hermeneutical injustice is non-agential and yet is a kind of epistemic injustice. Therefore, even if there is no agent of memorial injustice, that shouldn't prevent us from understanding memorial injustice as a species of epistemic injustice. However, this response is likely unsatisfactory, especially given that my motivation for pursuing this project proceeds from the analogy between memory and testimony. Testimonial injustice is agential (though structurally influenced), so we should expect that memorial injustice, *qua* analogue to testimonial injustice, is agential as well.

Expecting memorial injustice to be agential invites a further objection. Let us here return to the earlier question of whether remembering involves two selves or merely one. Earlier, I bracketed the question of personal identity and the metaphysics of the self over time, opting to use the terms "past self" and "present self" as shorthand for the self at the time of encoding and the self at the time of recall, without positing a particular ontological view of selfhood. The problem, then, is that in testimonial injustice, there are two agents: one who perpetrates the injustice and another who is victim to it. But if the ontological view of selfhood is that one remains the same self over time, then it is less clear how memorial injustice could be like testimonial injustice in this way. Though I myself remain agnostic on the ontology of selfhood, dodging this question might disappoint some readers' philosophical expectations. Here I suggest two different responses to the past self/present self worry, in the hope that the reader can adopt whichever view she prefers.

Response: Two Selves

Let us begin with the more straightforward—though, to my mind, less plausible—possibility, and suppose that the terms "past self" and "present self" are not mere shorthand for the self at the time of encoding and the self at the time of recall, respectively, but instead reflect real ontological commitments. Such a view aligns cleanly with Fricker's view of testimonial injustice as involving two agents, since it

implies similarly that two distinct agents are involved in a memorial exchange. The concern here, however, is that the past self is exactly that: past. There is no obvious way in which the past self suffers a harm, since her existence is either completely in the past or is otherwise inaccessible, and in either case, unlikely to be morally salient.

There are two things to be said in response: first, even inaccessible selves might bear moral standing. Consider, for instance, the moral standing of deceased persons. Though it is impossible to harm them in the conventional sense, most of us would still rankle at the thought of graverobbing, for instance, or vandalising a headstone, or failing to execute a will faithfully. Similarly, while the past self in an instance of remembering is no longer accessible, we should not take that to imply that he bears no moral standing whatsoever. He might still be a potential victim of harms, despite the fact that he might be unaware of them, and indeed, not consciously affected by them.

The other response is that the past self's stake in the situation is not a moral one. For Fricker, the primary harm is necessarily *epistemic*, and while moral or practical harms often follow downstream, they are contingent. Even if one thinks it is impossible to harm a dead person morally, it is important to bear in mind that considerations of moral harms do not necessarily translate to epistemic harms. In other words, it may well be impossible to harm a dead person *morally* and at the same time possible to harm a dead person *epistemically*. For instance, the diary of Anne Frank provides plenty of testimony as to the conditions of life in hiding in Nazi Germany. Even if we conceded, for the sake of argument, that she could no longer be harmed morally, this alone should not lead us to conclude that she could no longer be harmed epistemically. To reject the testimony recorded in Anne Frank's diary because of a credibility deficit resulting from prejudices about her age, gender, or religion does not seem to be different in any principled way from rejecting exactly the same testimony from a similar testifier living under the same conditions today. The only potentially relevant difference may be the inability to seek further

information from her, but as I remarked in Chapter 1, Section 1.7, this does not affect the diary's status as an instance of testimony.

Moreover, it is important to note that even if the epistemic harm to a past person is minor, there are still other potentially harmful downstream consequences of epistemic injustice, such as depriving others of important information. The most obvious 'other' in this case is the present self, who, like the people at the employment office in Carmita Wood's case, fail to learn what sexual harassment is or what Wood's experiences have been; they remain ignorant and fail to gain the relevant knowledge. The difference between the present self and the office employees who don't know what sexual harassment is, however, do not bear the same moral harms: the office employees come out morally unscathed, while the present self suffers a blow to self-confidence or self-respect as a result of not being able to access knowledge that is hers by right. In this way, the present self internalises and deploys prejudices such that she renders her past self the victim and present self the perpetrator of memorial injustice, while consequently depriving the present self both of the epistemic resources needed to understand herself and of her deserved corroboration of moral worth.

So, if we take the view that the past and present selves are real, we see that it is still possible to perpetrate an epistemic injustice because it is still possible to harm the past self epistemically. All one needs to do is not afford appropriate credibility to the past self; if that credibility deficit results from an identity prejudice, then the rejection of one's memories can amount to memorial injustice. Thus, there can still be an injustice even if it's *only* epistemic, i.e. even in the absence of a practical harm.

Moreover, there does seem to be a kind of practical harm. Whatever view one adopts about selfhood, it seems that even if there are two metaphysically distinct selves, the phenomenology of identity aligns with the intuition that there is only one self over time. At a minimum, past and present self seem to be deeply connected.

Thus, epistemically harming my past self might have psychological consequences for my present self, even if the two are ontologically distinct.

Response: One Self

Suppose, however, that you claim that the past self and present self are just one person. If this is the case, then the link to testimonial injustice is less clear, but the story is more plausible. All that would be required for a memorial injustice would be the rejection of memories owing to an identity-prejudicial credibility deficit. At first glance, this sounds both familiar and peculiar. It is familiar because, as we have seen, this formulation appears time and time again throughout the epistemic injustice literature. What makes this usage peculiar is that, if there is only one self, then an agent must deploy an identity prejudice against a group of which she is a member. For memorial injustice still to obtain, even if there is only one self, it must be the case that a person can do so.

The Matilda Effect (Rossiter 1993)—where the accomplishments of women are attributed to their male colleagues, such as Rosalind Franklin's contributions to the discovery of the double-helix structure of DNA, most commonly attributed only to James Watson and Francis Crick—serves as one example. Knobloch, Glynn, and Huge (2013) found that the preferential ranking of male-sounding and female-sounding scientist names did not vary along respondent sex; that is, female participants in the study exhibited the same preference for male-sounding scientist names as did their male counterparts.

The evidence goes beyond this study: gay men can internalise homophobia (Malyon 1982; Meyer 2003; Meyer and Dean 1998), women can internalise sexism (Bartky 1998; Bearman, Korobov, and Thorne 2009), people of colour can internalise racism (Speight 2007), and so on. Intellectual self-trust—which I take to be roughly equivalent to epistemic self-trust—is political (Jones 2012), and can be bolstered or threatened by social phenomena such as implicit bias and stereotype threat. So, it is

entirely plausible that a person could hold and deploy a prejudice against themself, thus giving rise to identity-prejudicial credibility deficits. Indeed, as Wright (forthcoming) notes, people can harm themselves epistemically through self-overestimation or self-underestimation, and when these harms arise from identity prejudices, they rise to the level of injustice in Fricker's sense.

There is one final point I wish to make about memorial injustice *qua* species of epistemic injustice. I claimed in Section 4.2 that testimonial injustice provided the best framework for understanding cases such as the false confessor. So, the reader may wonder how distinct memory and testimony really are. The upshot, in other words, is that if one thinks that memory is essentially a form of testimony, but nonparadigmatic in that it is internal to a single agent, then one will likely conclude that memorial injustice is likewise a nonparadigmatic form of testimonial injustice but not an altogether distinct form of epistemic injustice. Whether memorial injustice is best understood as a nonparadigmatic subspecies of testimonial injustice or as its own, separate species of epistemic injustice is an important question. Ultimately, I remain noncommittal on the issue, but suggest that one's ontological view on whether remembering is best understood as involving one or two selves will be one significant consideration for deciding between the two possibilities. In any case, even if memorial injustice is best understood as a nonparadigmatic form of testimonial injustice, the work I have done in this chapter should be a significant step toward elucidating its unique features.

4.4.2 Harms of memorial injustice

Like all species of epistemic injustice, memorial injustice's central harm is an epistemic one. I take it that this has been well-established by now; the doxastic shift I mentioned above and the undermining of one's status as a knower are paradigmatic examples of the central epistemic harm of memorial injustice. In the false confession case in particular, Kassin et al. (2010) have pointed out—somewhat cynically but with evidence—that the purpose of police interrogations is not to find the truth but

instead to secure a confession and conviction. Yet, disguised as an epistemic project, interrogations precipitate in suspects such a severe distrust that it undermines their status as knowers.[20] In more general terms, when power imbalances and social circumstances conspire to produce profound self-distrust, subjects suffer the central harm of memorial injustice.

In addition to the epistemic harm done to the individual, there are harms that target the individual as well as the broader epistemic community. Memorial injustice, *qua* species of epistemic injustice, "causes deficits in self-trust" (Jones 2012, 246), but so too do deficits in self-trust bring about epistemic injustice. Jones identifies this cycle: "unjust social relations cause epistemic injustice, which undermines self-trust among the underprivileged; unjust social relations create excessive self-trust among the privileged, which perpetuates epistemic injustice, which further undermines the self-trust of the disadvantaged in a vicious feedback loop" (Jones 2012, 247). So, epistemic injustice is self-perpetuating. Likewise, memorial injustice serves to undermine the self-trust of underprivileged knowers and to maintain repeated epistemic injustices.

Sue Campbell offers a vivid example, which I reproduce in full here for its richness:

> "Consider the following sort of relationship between a woman and her partner. Fights frequently escalate into some form of abusive behavior, and these fights frequently start as discussions about what happened on a past occasion. You, as the abused partner, are challenged to give an account of yourself in the past, and this account is then challenged with hostility. What are the probable effects of repetitions of this kind of situation? You are not being allowed to give your own account of the past when it is important that you should

[20] I suspect that if suspects knew that the aim of the game was to secure a confession, whether or not the confession was true, they would be less susceptible to memory distrust and memorial injustice, for their status as epistemic agents would not be under threat in the same way. This question is an empirical one, however, and I do not pursue it here.

> succeed at doing so. You may be put into doubt about the reliability of your evidence for your beliefs about yourself in the past and your beliefs about yourself in the present; and there may be consequences for how you regard your memory evidence for other important beliefs about yourself. You may be put into doubt as to whether your desires were reasonable or self-deceived, whether your actions were warranted, and whether what is of significance to you, as evidenced by what you remember, really is of significance. In sum, you may be put into doubt about the reliability of your memory as a source of warrant for your beliefs, desires, actions, and values in serious repeated situations. In becoming unsure of your descriptions of the past, you will be unsure as to who bears responsibility for past acts. Your abilities to assign responsibility, take responsibility, and be seen as responsible may all be threatened by the progressive distrust of your own recollections. Borland's response to Beatrice[21] displays an understanding of how someone can be undermined through a challenge to their memory, and it displays an attempt to avoid this undermining. Although it is not always possible to avoid undermining others, we can be on the lookout for strategies that deliberately weaken someone's ability to make sense of her or his past and his or her ability to negotiate responsibility for past acts. I believe that a view of personhood that is sensitive to the ways in which core cognitive abilities can be undermined can and must be used to locate the ways in which downward psychological constituting is implicated in abusive situations." (1997, 63-64)

Here, Campbell vividly illustrates the harms to the victim of memorial injustice, showing how memorial injustice loops back on itself and compounds its own effects over and over. The harms of memorial injustice to the individual are pernicious and repugnant, undermining an agent's self-trust in a way that other species of epistemic injustice do not seem to approach.

As bad as memorial injustice is for the individual, it also damages the collective pool of knowledge: as Hyde remarks in the testimonial injustice case, "the inappropriate [credibility] judgments caused by [metacognitive] failures may

[21] This is an example given earlier in the article, but as the details are not important here, I omit a summary.

prevent hearers' efforts at getting at the truth and further thwart speakers' rightful participation in the epistemic community" (2016, 864). When speakers cannot contribute satisfactorily to the body of knowledge, the entire community suffers through losing access to valuable knowledge. Similarly in the memorial injustice case, subjects who distrust their own memories so thoroughly that they reject them altogether pre-emptively silence themselves; the corresponding testimony never has the chance to make it into the collective pool of knowledge, for the content is disregarded before the testimony is uttered.

The harms of memorial injustice do not begin and end with the epistemic harms. Again, let us return to the false confession case, where the consequences are pronounced. One obvious practical harm is that a false confessor is likely to be tried and convicted of a crime he did not commit. All of the ordinary harms of wrongful conviction attend: incarceration, difficulty finding employment after incarceration, separation from loved ones, and so on. Even in memorial injustice cases aside from false confession, there might be practical harms, though without a concrete example it is difficult to spell them out in a meaningful way. In any case, as Fricker argues, the practical harms are contingent, and even if a given instance of memorial injustice bears no practical harm, it is no less an instance of memorial injustice than one that does.

The point is that, like other species of epistemic injustice, memorial injustice brings with it a variety of harms, both epistemic and practical, that affect the victim and the epistemic community at large.

Before concluding this chapter, I want to address a final objection to which I alluded above. So far, I have claimed that memorial injustice is internal to the remembering subject. But some readers may have noticed that my prime example, the false confessor, is not alone in the interrogation room. They may have noticed that his memorial injustice takes place in a testimonial setting, in which he offers testimony about his whereabouts and that testimony is rejected on ethically noxious

grounds. After repeated questioning, he gradually comes to reject his own memories and fill in the remaining void using information from the interrogators. But it is hard to imagine a circumstance in which a subject would spontaneously reject his own memories. You might therefore think that the false confessor is in the same boat as the laypersons who accept Bill Nye's testimony about the value of philosophy: that the false confessor has unjustly attributed a credibility excess to the interrogators and consequently accepts their testimony about his whereabouts on the night of the crime. If this is right, then the false confession is merely a case of testimonial injustice owing to identity-prejudicial credibility *excess* rather than *deficit*.

The response to this objection is twofold. First, while the interrogation is apparently an atypical situation, it is probably better understood as a concentrated version of ordinary life. Black men, for example, are routinely regarded and treated as dishonest and prone to crime or violence, the micro- and macroaggressions they experience in this vein will add up and erode their own epistemic self-trust over time. Women who constantly experience jokes about women's supposed ineptitude in mathematics, for example, may eventually genuinely come to believe that they are bad at math, even in the face of decisive evidence to the contrary. Challenges to one's reliability in a given domain can undermine one's confidence in that domain, endangering their epistemic self-trust in exactly the same way as the false confessor. We might, then, think of the interrogation room as a pressure cooker that throws into sharp relief what happens when one is routinely challenged as a knower. Gudjonsson and Lister (1984) found that individuals were more susceptible to influence from interrogators when they perceived greater disparity between themselves and the interrogators with respect to "competence, power, and control" (99), and that subjects' suggestibility and their perceived distance to the experimenters were highly significantly correlated. These findings lend further support to the notion that members of groups that are routinely challenged in their epistemic proficiency will be more likely to fall prey to memorial injustice, for it is

these individuals who are most likely to perceive a greater power disparity between themselves and interrogators. Indeed, as Gudjonsson et al. (2014) note, "most cases of false confession involve memory distrust being induced by police during prolonged and persuasive interviews, emphasising the importance of social influence" (338) to false confession. The importance of social influence cannot be overstated, and is what makes false confessions a striking example of epistemic injustice.

The second response to this objection is that even if we were to concede, for the sake of argument, that the false confession is just an ordinary testimonial injustice, we would be left unable to account for the corresponding doxastic shift in the agent. We cannot explain how the testimony of the interrogators can override the very memories of the suspect without, at a minimum, appealing to an internalised credibility deficit, which testimonial injustice alone cannot make sense of. Either way, we will need to import some additional conceptual resources to explain what is happening, and what I have presented in this chapter seeks to make plain the mechanisms to which we must appeal to explain memorial injustice.

So, while it is true that the suspect is not the only person in the room, and that the memorial injustice occurs under social circumstances, the important thing is that an internal doxastic shift takes place in a way that appears to be unique to memorial injustice. In other words, to call memorial injustice "purely internal" has been a convenient but perhaps misleading shorthand. The more precise way of putting it is to say that while the relevant epistemic phenomenon is internal, it takes place under certain external social conditions, and is thus not simply a matter of an agent perpetrating an injustice against himself. Instead, like all injustices, it is the result of a confluence of events, circumstances, agents, and most importantly, disparities in power.

4.5. Conclusion

In *Epistemic Injustice*, Fricker (2007, 44) writes that "we are long familiar with the idea, played out by the history of philosophy, that our rationality is what lends humanity its distinctive value. No wonder, then, that being insulted, undermined, or otherwise wronged in one's capacity as a giver of knowledge is something that can cut deep." I agree, but think that she does not go far enough. Fricker's focus on agents as *givers* of knowledge is, to my mind, well-justified, but as we have seen, epistemic injustice can cut still deeper than that. Memorial injustice does not merely undermine one's status as a giver of knowledge: it undermines one's status as a possessor of knowledge in the first place. Memorial injustice jeopardises one's potential to share knowledge because it jeopardises one's potential to have knowledge. It is therefore every bit as insidious as its cousin species, and brings the added threat of destabilising the internal epistemic lives of those it affects.

I conclude this chapter with a brief remark that I will revisit in the conclusion of this thesis. I have shown in this chapter that the way we evaluate our own memories and make judgments as to whether to accept or reject them is heavily influenced by social factors. Literature in social epistemology and moral psychology has already begun pointing this way: Bartky's (1990) chapter "Shame and Gender", Jones' (2012) "The Politics of Intellectual Self-Trust," and Dillon's (1997) "Self-Respect," to name a few, discuss how our conceptions of ourselves as full epistemic agents are politically and socially coloured. To the extent that our memories play vital roles in these self-conceptions, as well as how they are informed by them, social epistemologists should be thinking of memory in similar ways. Perhaps memory, like testimony, is better understood not as an individual enterprise, endeavour, capacity, or faculty, but as a social one.[22]

[22] I am grateful to audiences at the University of Otago philosophy postgraduate seminar and the 2019 New Zealand Association of Philosophers' conference and the New Zealand Association of Philosophers' 2019 Annual Conference for comments and discussion that improved this chapter.

CONCLUSION

I began this thesis with the relatively uncontroversial claim that memory and testimony alike are vital to our successful functioning as epistemic agents. Throughout the thesis, I have explored the various ways that memory and testimony serve as sources of knowledge, where remembering, roughly understood, is a matter of learning from oneself, and receiving testimony is a matter of learning from others.

In Chapter 1, I argued that memory and testimony are analogous, and that whatever epistemologists conclude about one, they must also conclude analogously about the other. Though I am not the first to advance this argument, the canonical version of it (Dummett 1994) takes memory to be a fundamentally transmissive capacity, transmitting beliefs from the past self to the present self. Until recently, epistemologists have taken this view of memory to be basically correct, despite the fact that the transmissive view is incompatible with empirical findings about memory (Michaelian 2011c). Instead, mounting research shows that memory is not a matter of transmission, but of reconstruction. I capitalised on memory's constructive features to defend the view that there is nevertheless an analogy between memory and testimony, and identified the corresponding constructive features of testimony.

The rest of the dissertation was devoted to identifying and exploring the implications of this analogy. In other words, I took as my starting point that memory and testimony are analogous, and determined what followed from there. I also worked to identify the ways in which they interact directly with each other; that is, I explored how testimony about an event influences a subject's memory of it. I also explored how understanding memory as a non-paradigmatic kind of testimony allows us to expand the notion of epistemic injustice (Fricker 2007).

In Chapter 2, I considered an implication of viewing testimony and memory as analogous. I proceeded from the observation that it is widely accepted that there is such a thing as memory-how, yet no testimonial counterpart exists; on the face of

it, this poses a problem for the memory-testimony analogy I defended in Chapter 1. For that reason, I set aside the topic of memory for a time and delved into the nature of testimony-how, i.e. exactly how it is that we might conceptualise instructions as a kind of non-paradigmatic testimony. I set up the problem as an inconsistent triad: (i) Instructions are non-propositional, (ii) instructions are a kind of testimony, and (iii) testimony is propositional. Ultimately, I argued that rejecting (i) or (ii) would not suffice as a solution to the problem, but that rejecting (iii) would. Testimony, I concluded, is not necessarily propositional. Although I did not explore this idea in depth, I suggested that there might be a way to posit an analogy between testimony-how and memory-how in much the same way that I had argued there was an analogy between testimony-that and memory-that in Chapter 1. Crudely put, my suggestion was that we can consider procedural memory to be a sort of testimony-how from the past self. This seems to me to be a promising avenue for future research.

In Chapter 3, I argued that if collective memory was like individual memory in the relevant ways, and if memory errors are a feature of individual memory, then it follows that memory errors are a feature of collective memory. The existence of collective memories is confirmed by the existence of the Mandela Effect (u/Jhoobie 2017) , which I argued was an example of collective confabulation. I argued that if one adopts a reliability account (Michaelian 2016a; 2016b) of memory, then the collective memories that constitute the Mandela Effect ("ME-memories") are confabulatory because they are produced by malfunctioning memory systems. On the other hand, if one adopts a causal account (Bernecker 2010; 2017; Robins 2016b; 2017) of memory, then ME-memories are confabulatory because they lack the appropriate causal connection to the past event. Whichever account one adopts, ME-memories are collective because interaction among members promotes the emergence of qualitative details through in-group testimony, such that the group

remembers more than does anyone individual. In short, ME-memories are collective confabulation.

Finally, in Chapter 4, I argued that there was such a thing as memorial injustice. First, I rehearsed the central argument of epistemic injustice, and particularly testimonial injustice, put forth by Miranda Fricker (2007). I argued that if the memory-testimony analogy holds, and if there is such a thing as testimonial injustice, then there is also be such a thing as memorial injustice, which might be well exemplified by internalised false confessions (Kassin and Wrightsman 1985). I considered how hostile and guilt-presumptive interrogations provide background conditions conducive to inducing memory distrust in innocent suspects, and compared these high-pressure situations to lower-pressure but equally ethically bad situations outside the interrogation room. I drew on literature about the role of mindreading in determining the trustworthiness of testifiers and compared it to the role of metacognition in memory, and concluded, following Hyde (2016), I argued that if testimonial injustice involves a failure of mindreading, then memorial injustice involves a failure of metacognition. Finally, I concluded with a sketch of an account of memorial injustice to join the growing family of species of epistemic injustice.

At this point, I think it is worth discussing how my arguments might provide the groundwork for future research. In Chapter 1, I was careful to emphasise that while I thought there was an epistemological and epistemic analogy between memory and testimony, I was agnostic with respect to what the epistemologies of the two really were. (For example, I took no stand in the reductionism/anti-reductionism debate.) Instead, I claimed simply that whatever epistemologists said about one, they would have to say (*mutatis mutandis*) about the other. Obviously, then, the natural progression is to investigate the epistemologies themselves. What *should* epistemologists say about memory and testimony? There has been plenty of discussion about this, of course, but as far as I can tell, almost none of it has taken

place under the presupposition that memory and testimony are analogous. One impact thinking about the two analogously might have is that if there is a compelling reason to reject a view (say, reductionism) in testimony, then that alone might give us reason to reject the analogous view in memory. We need not necessarily come up with independent reasons to reject reductionism in memory. If the analogy holds, it provides a different set of parameters for us to pursue the epistemologies of memory and testimony.

The focus in Chapter 2 was on the non-propositionality of testimony, which I pursued with the hope that showing that testimony could be non-propositional would create space to talk about an analogy between testimony-how and memory-how. The obvious next step is to marry these claims with psychological research about the nature of procedural memory to see whether the broad memory-testimony analogy holds as tightly as I think it will. Indeed, the claims I have made have been in accordance with the generally accepted taxonomy of memory, but should that taxonomy ever come under serious criticism, it would certainly be worth exploring what that means for the analogy I have proposed here.

In Chapter 3, I argued that the Mandela Effect amounted to a type of collective confabulation. While I think the Mandela Effect provides a neatly contained, slightly humourous example, the phenomena and mechanisms that give rise to it can also give rise to much more significant collective confabulations. In the political sphere, collective confabulation can contribute to the growing divide between social groups. As has been noted by countless others, political debates are no longer about what the best responses to the facts are; they are about the facts themselves. They are disagreements about what is true and what really happened, not merely about economic policy or electoral reform or tax brackets. When we take seriously how testimony from others is incorporated into memory on a group scale, particularly in echo chambers, we can begin to see how collective confabulation compounds and feeds into the epistemic schism. I think it would be worth

investigating collective confabulation and its relation to politics and political discourse.

Throughout Chapter 4, I relied on the example of internalised false confession to make plain some of the features of memorial injustice. However, memorial injustice has broader reach and might provide conceptual resources that go some way toward explaining several other epistemically specious phenomena. For instance, gaslighting is a kind of psychological abuse in which an abuser seeks to manipulate a victim into agreeing with him by gradually eroding her epistemic self-trust (Abramson 2014; Calef and Winshel 1981; Spear 2020). Central to successful gaslighting is the abuser's reliance on "manipulation, fabrication, and deception" (Spear 2020, 230) to force the deterioration of the victim's self-trust. Gaslighting typically depends on the victim's perception of the gaslighter as an epistemic peer or authority. The points of contact with internalised false confession should be clear: an imbalance of epistemic and social power erodes the victim's self-trust such that she eventually experiences a doxastic shift consonant with the abuser's way of seeing the world (or at least the abuser's desired way for the victim to see the world). Memorial injustice can provide a way toward understanding the distinctly epistemically unjust dimensions of gaslighting.

Memorial injustice, I suspect, can also do something to explain phenomena such as impostor syndrome (Clance and Imes 1978) or the Dunning-Kruger Effect (Dunning 2011), where the two are understood as deflated and inflated self-directed credibility judgments, respectively. If persistent challenges to one's competence in a domain gradually erode the subject's epistemic self-trust, as is the case in impostor syndrome, or unearned epistemic overconfidence artificially inflates one's judgments of oneself, as is the case in the Dunning-Kruger Effect, such that the relevant doxastic shifts take place, then it appears as though something much like internalised false confessions is going on. I think that with further research,

memorial injustice (or at least the mechanisms that underpin it) can provide a fruitful way to think about these issues.

I will offer one final remark on the twin roles that memory and testimony play in our epistemic lives. Social epistemology takes as its starting point the claim that knowledge is best understood not in the heavily individualistic way that abstracts knowers from their social locations, but as a feature of human existence that is heavily shaped by social interactions, norms, and institutions. The obvious target for social epistemologists is testimony, which is steeped in social mores and heavily influenced by power relations, assessments of authority and expertise, and prejudice. Testimony can be adversarial or collaborative, constructive or destructive—inevitably, social interactions and institutions inform the norms of testimony, how it is produced, and how it is taken up. For this reason, testimony has long been the bread and butter of social epistemologists.

But if what I have said in this thesis is correct—and naturally, I am inclined to think it is—and if memory is not so different from testimony after all, and if the two interact with each other in ways that alter the beliefs each produces (the Mandela Effect being a case in point), then it is high time that social epistemologists turned their attention to the way that memory is knit into the same social fabric as testimony. Memory has long occupied a position on the pedestal of basic sources of knowledge, right alongside perception and reason. If we attend to the myriad ways in which social phenomena influence memory—whether that be through testimony integrated into memories, through the culturally specific ways of remembering, or through our credence in our own memories depending on our social location, just to name a few—we begin to see that memory, even garden-variety biological individual memory, is anything but the purely internal and individual process it has long been taken to be. I hope that this thesis can serve as a starting point for further steps toward understanding what this really means for us, not only as epistemologists, but as knowers.

REFERENCES

Abramson, Kate. 2014. "Turning up the Lights on Gaslighting." *Philosophical Perspectives* 28 (1): 1–30. https://doi.org/10.1111/phpe.12046.

Adler, Jonathan. 2017. "Epistemological Problems of Testimony." In *Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. https://plato.stanford.edu/archives/win2017/entries/testimony-episprob/.

Alba, Joseph W., and Lynn Hasher. 1983. "Is Memory Schematic?" *Psychological Bulletin* 93 (2): 203–31.

Anderson, A.R., and O.K. Moore. 1957. "The Formal Analysis of Normative Concepts." *American Sociological Review* 22 (1): 9–17.

Anderson, Derek Egan. 2017. "Conceptual Competence Injustice." *Social Epistemology* 31 (2): 210–23. https://doi.org/10.1080/02691728.2016.1241320.

Anderson, Elizabeth. 2012. "Epistemic Justice as a Virtue of Social Institutions." *Social Epistemology* 26 (2): 163–73. https://doi.org/10.1080/02691728.2011.652211.

Audi, Robert. 1997. "The Place of Testimony in the Fabric of Knowledge and Justification." *American Philosophical Quarterly* 34 (4): 405–22.

———. 2006. "Testimony, Credulity, and Veracity." In *The Epistemology of Testimony*, edited by Jennifer Lackey and Ernest Sosa, 25–46. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199276011.003.0002.

———. 2013. "Testimony as a Social Foundation of Knowledge." *Philosophy and Phenomenological Research* 87 (3): 507–31. https://doi.org/10.1111/j.1933-1592.2011.00525.x.

Ballantyne, Nathan. 2019. "Epistemic Trespassing." *Mind* 128 (510): 367–95. https://doi.org/10.1093/mind/fzx042.

Barnier, Amanda J., John Sutton, Celia B. Harris, and Robert A. Wilson. 2008. "A Conceptual and Empirical Framework for the Social Distribution of Cognition: The Case of Memory." *Cognitive Systems Research* 9 (1–2): 33–51. https://doi.org/10.1016/j.cogsys.2007.07.002.

Bartky, Sandra Lee. 1990. *Femininity and Domination: Studies in the Phenomenology of Oppression*. Thinking Gender. New York: Routledge.

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

Bearman, Steve, Neill Korobov, and Avril Thorne. 2009. "The Fabric of Internalized Sexism" 1: 38.

Bengson, John, and Marc A. Moffett. 2011a. "Nonpropositional Intellectualism." In *Knowing How: Essays on Knowledge, Mind, and Action*, edited by John Bengson and Marc A. Moffett, 161–95. New York: Oxford University Press.

———. 2011b. "Two Conceptions of Mind and Action: Knowing How and the Philosophical Theory of Intelligence." In *Knowing How: Essays on Knowledge,*

Mind, and Action, edited by John Bengson and Marc A. Moffett, 3–58. New York: Oxford University Press.
Bergen, Saskia van. 2011. "Memory Distrust in Legal Context." University of Maastricht.
Bergen, Saskia van, Robert Horselenberg, Harald Merckelbach, Marko Jelicic, and Roos Beckers. 2010. "Memory Distrust and Acceptance of Misinformation." *Applied Cognitive Psychology* 24 (6): 885–96. https://doi.org/10.1002/acp.1595.
Berlyne, N. 1972. "Confabulation." *The British Journal of Psychiatry* 120 (554): 31–39. https://doi.org/10.1192/bjp.120.554.31.
Bernecker, Sven. 2008. *The Metaphysics of Memory*. Springer Science & Business Media.
———. 2010. *Memory: A Philosophical Study*. Oxford, New York: Oxford University Press.
———. 2017. "A Causal Theory of Mnemonic Confabulation." *Frontiers in Psychology* 8 (July). https://doi.org/10.3389/fpsyg.2017.01207.
Berntsen, Dorthe. 2010. "The Unbidden Past: Involuntary Autobiographical Memories as a Basic Mode of Remembering." *Current Directions in Psychological Science* 19 (3): 138–42. https://doi.org/10.1177/0963721410370301.
Blum, Lawrence. 2004. "Stereotypes And Stereotyping: A Moral Analysis." *Philosophical Papers* 33 (3): 251–89. https://doi.org/10.1080/05568640409485143.
Bohnert, H.G. 1945. "The Semiotic Status of Commands." *Philosophy of Science* 12 (4): 302–15.
Bond, Charles F., and Bella M. DePaulo. 2006. "Accuracy of Deception Judgments." *Personality and Social Psychology Review* 10 (3): 214–34. https://doi.org/10.1207/s15327957pspr1003_2.
Bortolotti, Lisa, and Rochelle E. Cox. 2009. "'Faultless' Ignorance: Strengths and Limitations of Epistemic Definitions of Confabulation." *Consciousness and Cognition* 18 (4): 952–65. https://doi.org/10.1016/j.concog.2009.08.011.
Brainerd, C.J., and V.F. Reyna. 2005. *The Science of False Memory*. Oxford: Oxford University Press.
Brewer, William F., and James C. Treyens. 1981. "Role of Schemata in Memory for Places." *Cognitive Psychology* 13 (2): 207–230.
Brockmeier, Jens. 2015. *Beyond the Archive: Memory, Narrative, and the Autobiographical Process*. Oxford: Oxford University Press.
Broome, Fiona. 2009. "Mandela Effect - Alternate Realities." The Mandela Effect. 2009. https://mandelaeffect.com/.
———. 2013. "Billy Graham's Funeral on TV." *Mandela Effect* (blog). April 25, 2013. https://mandelaeffect.com/billy-grahams-funeral-on-tv/.
———. 2016. "DiCaprio Wins... Again?" *Mandela Effect* (blog). March 1, 2016. https://mandelaeffect.com/dicaprio-wins-again/.
Brown, Alan S., Kathryn Croft Caderao, Lindy M. Fields, and Elizabeth J. Marsh. 2015. "Borrowing Personal Memories: Borrowing Personal Memories." *Applied Cognitive Psychology* 29 (3): 471–77. https://doi.org/10.1002/acp.3130.

Brown, D. G. 1974. “Reply to Brett.” *Canadian Journal of Philosophy* 4 (2): 301–3. https://doi.org/10.1080/00455091.1974.10716940.

Brown, D.G. 1971. “Knowing How and Knowing That, What.” In *Ryle: A Collection of Critical Essays*, edited by Oscar P. Wood and George Pitcher, 213–48. Garden City, NY: Anchor Books.

Buckner, Randy L., and Daniel C. Carroll. 2007. “Self-Projection and the Brain.” *Trends in Cognitive Sciences* 11 (2): 49–57. https://doi.org/10.1016/j.tics.2006.11.004.

Burge, Tyler. 1993. “Content Preservation.” *The Philosophical Review* 102 (4): 457. https://doi.org/10.2307/2185680.

Campbell, Sue. 1997. “Women, ‘False’ Memory, and Personal Identity.” *Hypatia* 12 (2): 51–82.

———. 2003. *Relational Remembering: Rethinking the Memory Wars*. Lanham, Md: Rowman & Littlefield Publishers.

Cappelen, Herman, and Ernie LePore. 2005. *Insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism*. Malden, MA: Blackwell Pub.

Carr, David. 1979. “The Logic of Knowing How and Ability.” *Mind* 88: 394–409.

Carruthers, Peter. 2006. *The Architecture of the Mind*. Oxford University Press. http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199207077.001.0001/acprof-9780199207077.

———. 2009. “How We Know Our Own Minds: The Relationship between Mindreading and Metacognition.” *Behavioral and Brain Sciences* 32 (02): 121–38. https://doi.org/10.1017/S0140525X09000545.

Carston, Robyn. 2002. *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell.

Carter, J. Adam, and Duncan Pritchard. 2015. “Knowledge-How and Epistemic Luck.” *Noûs* 49 (3): 440–53. https://doi.org/10.1111/nous.12054.

Chakrabarti, Arindam. 1994. “Telling as Letting Know.” In *Knowing From Words: Western and Indian Philosophical Analysis of Understanding and Testimony*, edited by Bimal K. Matilal and Arindam Chakrabarti, 99–124. Springer.

Chapman, Frances E. 2013. “Coerced Internalized False Confessions and Police Interrogations: The Power of Coercion.” *Law & Psychology Review; Tuscaloosa* 37: 159–92.

Churchland, Paul. 1988. *Matter and Consciousness*. Cambridge, Mass.: MIT Press.

Clance, Pauline Rose, and Suzanne Ament Imes. 1978. “The Imposter Phenomenon in High Achieving Women: Dynamics and Therapeutic Intervention.” *Psychotherapy: Theory, Research & Practice* 15 (3): 241–47. https://doi.org/10.1037/h0086006.

Clark-Younger, Hannah. 2014. “Imperatives and Logical Consequence.” Thesis, Dunedin, New Zealand: University of Otago. https://ourarchive.otago.ac.nz/handle/10523/5039.

Coady, C.A.J. 1992. *Testimony: A Philosophical Study*. New York: Clarendon Press.

Code, Lorraine B. 1981. “Is the Sex of the Knower Epistemologically Significant?” *Metaphilosophy* 12 (3–4): 267–76. https://doi.org/10.1111/j.1467-9973.1981.tb00760.x.

Collins, Patricia Hill. 1986. “Learning from the Outsider Within: The Sociological Significance of Black Feminist Thought.” *Social Problems* 33 (6): 14–32.

———. 1990. *Black Feminist Thought: Knowledge, Consciousness and the Politics of Empowerment*. New York: Routledge.

De Brigard, Felipe. 2014. “Is Memory for Remembering? Recollection as a Form of Episodic Hypothetical Thinking.” *Synthese* 191 (2): 155–85. https://doi.org/10.1007/s11229-013-0247-7.

Deese, James. 1959. “On the Prediction of Occurrence of Particular Verbal Intrusions in Immediate Recall.” *Journal of Experimental Psychology* 58 (1): 17.

Devitt, Aleea L., Edwin Monk-Fromont, Daniel L. Schacter, and Donna Rose Addis. 2016. “Factors That Influence the Generation of Autobiographical Memory Conjunction Errors.” *Memory* 24 (2): 204–22. https://doi.org/10.1080/09658211.2014.998680.

Dillon, Robin S. 1997. “Self-Respect: Moral, Emotional, Political.” *Ethics* 107 (2): 226–49. https://doi.org/10.1086/233719.

Dotson, Kristie. 2011. “Tracking Epistemic Violence, Tracking Practices of Silencing.” *Hypatia* 26 (2): 236–57. https://doi.org/10.1111/j.1527-2001.2011.01177.x.

———. 2012. “A Cautionary Tale: On Limiting Epistemic Oppression.” *Frontiers: A Journal of Women Studies* 33 (1): 24. https://doi.org/10.5250/fronjwomestud.33.1.0024.

Dubois, Elizabeth, and Grant Blank. 2018. “The Echo Chamber Is Overstated: The Moderating Effect of Political Interest and Diverse Media.” *Information, Communication & Society* 21 (5): 729–45. https://doi.org/10.1080/1369118X.2018.1428656.

Dummett, Michael. 1994. “Testimony and Memory.” In *Knowing From Words: Western and Indian Philosophical Analysis of Understanding and Testimony*, edited by Bimal K. Matilal and A. Chakrabarti, 251–272. Springer.

Dunning, David. 2011. “The Dunning–Kruger Effect.” In *Advances in Experimental Social Psychology*, 44:247–96. Elsevier. https://doi.org/10.1016/B978-0-12-385522-0.00005-6.

Efklides, Anastasia. 2015. *Metamemory and Affect*. Edited by John Dunlosky and Sarah (Uma) K. Tauber. Vol. 1. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199336746.013.1.

Fantl, Jeremy. 2008. “Knowing-How and Knowing-That.” *Philosophy Compass* 3 (3): 451–70. https://doi.org/10.1111/j.1747-9991.2008.00137.x.

Feinberg, Todd E. 2001. *Altered Egos: How the Brain Creates the Self*. Oxford: Oxford University Press.

Fernández, Jordi. 2006. “The Intentionality of Memory.” *Australasian Journal of Philosophy* 84 (1): 39–57. https://doi.org/10.1080/00048400600571695.

———. 2018. “The Functional Character of Memory.” In *New Directions in the Philosophy of Memory*, edited by Kourken Michaelian, Dorothea Debus, and Denis Perrin, 52–72. Routledge Studies in Contemporary Philosophy. New York: Routledge.
Fodor, Jerry. 1987. *Psychosemantics*. Cambridge, Mass.: MIT Press.
Fricker, Elizabeth. 1994. “Against Gullibility.” In *Knowing From Words*, edited by Bimal K. Matilal and A. Chakrabarti, 125–61. Dordrecht: Kluwer.
———. 1995. “Telling and Trusting: Reductionism and Anti-Reductionism in the Epistemology of Testimony.” *Mind* 104 (414): 393–411.
Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.
Gallese, Vittorio. 2001. “The ‘“Shared Manifold”’ Hypothesis: From Mirror Neurons to Empathy.” *Journal of Consciousness Studies* 8 (5–7): 33–50.
Gallo, David A., and James M. Lampinen. 2015. *Three Pillars of False Memory Prevention*. Edited by John Dunlosky and Sarah (Uma) K. Tauber. Vol. 1. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199336746.013.11.
Garrett, R. Kelly. 2017. “The ‘Echo Chamber’ Distraction: Disinformation Campaigns Are the Problem, Not Audience Fragmentation.” *Journal of Applied Research in Memory and Cognition* 6 (4): 370–76. https://doi.org/10.1016/j.jarmac.2017.09.011.
Gazzaniga, Michael S. 2009. “Consciousness and the Cerebral Hemispheres.” In *The Cognitive Neurosciences*, 4th ed. Cambridge, Mass.: MIT Press.
Gelfert, Axel. 2014. *A Critical Introduction to Testimony*. London: Bloomsbury.
Gendler, Tamar Szabó. 2011. “On the Epistemic Costs of Implicit Bias.” *Philosophical Studies* 156 (1): 33–63. https://doi.org/10.1007/s11098-011-9801-7.
Gibbons, P. C. 1960. “Imperatives and Indicatives (I).” *Australasian Journal of Philosophy* 38 (2): 107–19. https://doi.org/10.1080/00048406085200121.
Ginet, Carl. 1975. *Knowledge, Perception, and Memory*. Dordrecht: D. Reidel Publishing Company.
Goldberg, Sanford. 2005. “Testimonial Knowledge through Unsafe Testimony.” *Analysis* 65 (4): 302–11.
Goldberg, Sanford C. 2001. “Testimonially Based Knowledge from False Testimony.” *The Philosophical Quarterly* 51 (205): 512–26. https://doi.org/10.1111/1467-9213.00244.
Goldman, Alvin. 1992. “In Defense of the Simulation Theory.” *Mind & Language* 7 (1–2): 104–19. https://doi.org/10.1111/j.1468-0017.1992.tb00200.x.
———. 2005. “Imitation, Mind Reading, and Simulation.” In *Perspectives on Imitation: Imitation, Human Development, and Culture*, edited by Susan Hurley and Nick Chater, 2:79–94. Cambridge, Mass.: MIT Press. https://mitpress.mit.edu/books/perspectives-imitation-volume-2.

Gopnik, Alison. 1993. "How We Know Our Minds: The Illusion of First-Person Knowledge of Intentionality." *Behavioral and Brain Sciences* 16 (March): 1–14. https://doi.org/10.1017/S0140525X00028636.
Gordon, Robert M. 1986. "Folk Psychology as Simulation." *Mind & Language* 1 (2): 158–171.
Graham, Peter J. 1997. "What Is Testimony?" *The Philosophical Quarterly* 47 (187): 227–32. https://doi.org/10.1111/1467-9213.00057.
———. 2000. "Conveying Information." *Synthese* 123: 365–92.
———. 2006. "Can Testimony Generate Knowledge?" *Philosophica*, no. 78: 105–27.
Green, Christopher R. 2006. "The Epistemic Parity of Testimony, Memory, and Perception." Notre Dame, Indiana: University of Notre Dame.
Grice, H.P. 1975. "Logic and Conversation." In *Syntax and Semantics*, edited by P. Cole and J.L. Morgan, 3:41–58. New York: Academic Press.
Gudjonsson, Gisli H. 2003. *The Psychology of Interrogations and Confessions: A Handbook*. John Wiley & Sons.
Gudjonsson, Gisli H., and S. Lister. 1984. "Interrogative Suggestibility and Its Relationship with Self-Esteem and Control." *Journal of the Forensic Science Society* 24 (2): 99–110. https://doi.org/10.1016/S0015-7368(84)72302-4.
Gudjonsson, Gisli H., and J.A.C. MacKeith. 1982. "False Confessions: Psychological Effects of Interrogation." In *Reconstructing the Past: The Role of Psychologists in Criminal Trials*, edited by Arne Trankell, 253–69. Deventer, The Netherlands: Kluwer.
Gudjonsson, Gisli H., Jon Fridrik Sigurdsson, Arndis Soffia Sigurdardottir, Haraldur Steinthorsson, and Valgerdur Maria Sigurdardottir. 2014. "The Role of Memory Distrust in Cases of Internalised False Confession." *Applied Cognitive Psychology* 28 (3): 336–48. https://doi.org/10.1002/acp.3002.
Habgood-Coote, Joshua. 2018a. "Knowledge-How, Abilities, and Questions." *Australasian Journal of Philosophy*, February, 1–19. https://doi.org/10.1080/00048402.2018.1434550.
———. 2018b. "Knowing-How, Showing, and Epistemic Norms." *Synthese* 195 (8): 3597–3620. https://doi.org/10.1007/s11229-017-1389-9.
Hakli, Raul. 2006. "Group Beliefs and the Distinction between Belief and Acceptance." *Cognitive Systems Research* 7 (2–3): 286–97. https://doi.org/10.1016/j.cogsys.2005.11.013.
———. 2007. "On the Possibility of Group Knowledge without Belief." *Social Epistemology* 21 (3): 249–66. https://doi.org/10.1080/02691720701685581.
Hamblin, C.L. 1987. *Imperatives*. Oxford: Basil Blackwell.
Harding, Sandra. 1991. *Whose Science? Whose Knowledge?* Ithaca, NY: Cornell University Press. http://www.jstor.org/stable/10.7591/j.ctt1hhfnmg.
———. 1992. "Rethinking Standpoint Epistemology: What Is 'Strong Objectivity?'" *The Centennial Review* 36 (3): 437–70.
Hare, R.M. 1952. *The Language of Morals*. Oxford: Clarendon Press.

Harris, Celia B., Amanda J. Barnier, John Sutton, and Paul G. Keil. 2014. "Couples as Socially Distributed Cognitive Systems: Remembering in Everyday Social and Material Contexts." *Memory Studies* 7 (3): 285–297.

Harris, Celia B., Amanda J. Barnier, John Sutton, Paul G. Keil, and Roger A. Dixon. 2017. "'Going Episodic': Collaborative Inhibition and Facilitation When Long-Married Couples Remember Together." *Memory* 25 (8): 1148–59. https://doi.org/10.1080/09658211.2016.1274405.

Harris, Paul L. 1989. *Children and Emotion*. Malden, MA: Blackwell Publishing.

Hartland-Swann, John. 1956. "'Knowing That'--A Reply to Mr. Ammerman." *Analysis* 17 (2): 69–71.

———. 1957. "The Logical Status of 'Knowing That.'" *Analysis* 16 (5): 111–15.

Hartsock, Nancy C.M. 1983. "The Feminist Standpoint: Developing the Ground for a Specifically Feminist Historical Materialism." In *Discovering Reality: Feminist Perspectives on Epistemology, Metaphysics, Methodology, and Philosophy of Science*, edited by Sandra Harding and Merrill B. Hintikka, 283–310. New York: Kluwer.

Hassabis, Demis, and Eleanor A. Maguire. 2007. "Deconstructing Episodic Memory with Construction." *Trends in Cognitive Sciences* 11 (7): 299–306. https://doi.org/10.1016/j.tics.2007.05.001.

Hawley, Katherine. 2003. "Success and Knowledge-How." *American Philosophical Quarterly* 40 (1): 19–31.

———. 2010. "Testimony and Knowing How." *Studies in History and Philosophy of Science Part A* 41 (4): 397–404. https://doi.org/10.1016/j.shpsa.2010.10.005.

Heal, Jane. 1986. "Replication and Functionalism." In *Language, Mind, and Logic*, edited by Jeremy Butterfield, 135–50. Cambridge: Cambridge University Press.

Henkel, Linda A., and Kimberly J. Coffman. 2004. "Memory Distortions in Coerced False Confessions: A Source Monitoring Framework Analysis." *Applied Cognitive Psychology* 18 (5): 567–88. https://doi.org/10.1002/acp.1026.

Hetherington, Stephen. 2006. "How to Know (That Knowledge-That Is Knowledge-How)." In *Epistemology Futures*, edited by Stephen Hetherington, 71–94. Oxford: Clarendon Press.

Hinchman, Edward S. 2005. "Telling as Inviting to Trust." *Philosophy and Phenomenological Research* 70 (3): 562–87. https://doi.org/10.1111/j.1933-1592.2005.tb00415.x.

Hirstein, William. 2005. *Brain Fiction: Self-Deception and the Riddle of Confabulation*. Philosophical Psychopathology. Cambridge, Mass: MIT Press.

Holroyd, Jules. 2012. "Responsibility for Implicit Bias." *Journal of Social Philosophy* 43 (3): 274–306. https://doi.org/10.1111/j.1467-9833.2012.01565.x.

Holroyd, Jules, Robin Scaife, and Tom Stafford. 2017. "Responsibility for Implicit Bias." *Philosophy Compass* 12 (3): e12410. https://doi.org/10.1111/phc3.12410.

hooks, bell. 1984. *Feminist Theory: From Margin to Center*. Boston: South End Press. https://doi.org/10.4324/9781315743172.

Huebner, Bryce. 2014. *Macrocognition: A Theory of Distributed Minds and Collective Intentionality*. Oxford University Press.
———. 2016. "Transactive Memory Reconstructed: Rethinking Wegner's Research Program." *The Southern Journal of Philosophy* 54 (1): 48–69. https://doi.org/10.1111/sjp.12160.
Hyde, Krista. 2016. "Testimonial Injustice and Mindreading." *Hypatia* 31 (4): 858–73. https://doi.org/10.1111/hypa.12273.
Ikier, Simay, Ali İ Tekcan, Sami Gülgöz, and Aylin C. Küntay. 2003. "Whose Life Is It Anyway? Adoption of Each Other's Autobiographical Memories by Twins." *Applied Cognitive Psychology* 17 (2): 237–47. https://doi.org/10.1002/acp.869.
Jack, Julie. 1994. "The Role of Comprehension." In *Knowing From Words: Western and Indian Philosophical Analysis of Understanding and Testimony*, edited by Bimal K. Matilal and Arindam Chakrabarti, 163–93. Springer.
Jackson, Frank, and Philip Pettit. 1998. "A Problem for Expressivism." *Analysis* 58 (4): 239–51.
Johnson, Marcia K., Shahin Hashtroudi, and D. Stephen Lindsay. 1993. "Source Monitoring." *Psychological Bulletin* 114 (1): 3–28.
Jones, Karen. 2012. "The Politics of Intellectual Self-Trust." *Social Epistemology* 26 (2): 237–51. https://doi.org/10.1080/02691728.2011.652215.
Jörgensen, Jörgen. 1937. "Imperatives and Logic." *Erkenntnis* 7: 288–96.
Kahan, Dan M. 2012. "Ideology, Motivated Reasoning, and Cognitive Reflection: An Experimental Study."
Kassin, Saul M. 2014. "False Confessions: Causes, Consequences, and Implications for Reform." *Policy Insights from the Behavioral and Brain Sciences* 1 (1): 112–21. https://doi.org/10.1177/2372732214548678.
Kassin, Saul M., Steven A. Drizin, Thomas Grisso, Gisli H. Gudjonsson, Richard A. Leo, and Allison D. Redlich. 2010. "Police-Induced Confessions: Risk Factors and Recommendations." *Law and Human Behavior* 34 (1): 3–38.
Kassin, Saul M., and Lawrence S. Wrightsman. 1985. "Confession Evidence." In *The Psychology of Evidence and Trial Procedure*, edited by Saul M. Kassin and Lawrence S. Wrightsman, 67–94. Beverley Hills: Sage Publications. https://web.williams.edu/Psychology/Faculty/Kassin/files/kassin_wrightsman_1985.pdf.
Katzoff, Charlotte. 1984. "Knowing How." *Southern Journal of Philosophy* 22: 61–67.
Kaufmann, Magdalena. 2012. *Interpreting Imperatives*. 1st ed. Studies in Linguistics and Philosophy 88. New York: Springer.
Kenyon, Tim. 2013. "The Informational Richness of Testimonial Contexts." *The Philosophical Quarterly* 63 (250): 58–80. https://doi.org/10.1111/1467-9213.12000.
Knobloch-Westerwick, Silvia, Carroll J. Glynn, and Michael Huge. 2013. "The Matilda Effect in Science Communication: An Experiment on Gender Bias in Publication Quality Perceptions and Collaboration Interest." *Science Communication* 35 (5): 603–25. https://doi.org/10.1177/1075547012472684.

Kopelman, Michael D. 1999. “Varieties of False Memory.” *Cognitive Neuropsychology* 16 (3–5): 197–214. https://doi.org/10.1080/026432999380762.
Koriat, Asher, and Morris Goldsmith. 1996. “Memory Metaphors and the Real-Life/Laboratory Controversy: Correspondence versus Storehouse Conceptions of Memory.” *Behavioral and Brain Sciences* 19 (02): 167. https://doi.org/10.1017/S0140525X00042114.
Küntay, Aylïn C., Samï Gülgöz, and Alï İ Tekcan. 2004. “Disputed memories of twins: how ordinary are they?” *Applied Cognitive Psychology* 18 (4): 405–13. https://doi.org/10.1002/acp.976.
Lackey, Jennifer. 1999. “Testimonial Knowledge and Transmission.” *The Philosophical Quarterly* 49 (197): 471–90. https://doi.org/10.1111/1467-9213.00154.
———. 2005. “Memory as a Generative Epistemic Source.” *Philosophy and Phenomenological Research* 70 (3): 636–58.
———. 2006a. “Introduction.” In *The Epistemology of Testimony*, edited by Jennifer Lackey and Ernest Sosa, 1–21. Oxford: Oxford University Press.
———. 2006b. “Learning from Words.” *Philosophy and Phenomenological Research* 73 (1): 77–101.
———. 2006c. “Knowing from Testimony.” *Philosophy Compass* 1 (5): 432–48. https://doi.org/10.1111/j.1747-9991.2006.00035.x.
———. 2007. “Why Memory Really Is a Generative Epistemic Source: A Reply to Senor.” *Philosophy and Phenomenological Research* 74 (1): 209–19.
———. 2008. *Learning from Words: Testimony as a Source of Knowledge*. OUP Oxford.
———. 2015. “Reliability and Knowledge in the Epistemology of Testimony.” *Episteme* 12 (02): 203–8. https://doi.org/10.1017/epi.2015.26.
Lackey, Jennifer, and Ernest Sosa, eds. 2006. *The Epistemology of Testimony*. Oxford University Press.
Lane, Andrew M., Gregory P. Whyte, Peter C. Terry, and Alan M. Nevill. 2005. “Mood, Self-Set Goals and Examination Performance: The Moderating Effect of Depressed Mood.” *Personality and Individual Differences* 39 (1): 143–53. https://doi.org/10.1016/j.paid.2004.12.015.
Lehrer, Keith. 2006. “Testimony and Trustworthiness.” In *The Epistemology of Testimony*, edited by Jennifer Lackey and Ernest Sosa, 145–59. Oxford: Oxford University Press.
Lewandowsky, Stephan, Ullrich K.H. Ecker, and John Cook. 2017. “Beyond Misinformation: Understanding and Coping with the ‘Post-Truth’ Era.” *Journal of Applied Research in Memory and Cognition* 6 (4): 353–69. https://doi.org/10.1016/j.jarmac.2017.07.008.
Lewis, David. 1969. *Convention: A Philosophical Study*. Cambridge: Harvard University Press.
———. 1970. “General Semantics.” *Synthese* 22 (1–2): 18–67.
———. 1990. “What Experience Teaches.” In *Mind and Cognition*, edited by William G. Lycan, 469–77. Oxford: Blackwell.

Liebow, Nabina. 2016. “Internalized Oppression and Its Varied Moral Harms: Self-Perceptions of Reduced Agency and Criminality.” *Hypatia* 31 (4): 713–29. https://doi.org/10.1111/hypa.12265.

Lister, Richard G, Michael J. Eckardt, and Herbert Weingartner. 1987. “Ethanol Intoxication and Memory.” In *Recent Developments in Alcoholism*, edited by Marc Galanter, 5:111–26. New York: Springer Science & Business Media.

Loftus, Elizabeth F. 1997a. “Creating Childhood Memories.” *Applied Cognitive Psychology* 11 (7): S75–86. https://doi.org/10.1002/(SICI)1099-0720(199712)11:7<S75::AID-ACP514>3.0.CO;2-F.

———. 1997b. “Memory for a Past That Never Was.” *Current Directions in Psychological Science* 6 (3): 60–65.

———. 2005. “Planting Misinformation in the Human Mind: A 30-Year Investigation of the Malleability of Memory.” *Learning & Memory* 12 (4): 361–66. https://doi.org/10.1101/lm.94705.

Loftus, Elizabeth F, Helen J. Burns, and David G. Miller. 1978. “Semantic Integration of Verbal Information into a Visual Memory.” *Journal of Experimental Psychology: Human Learning and Memory* 4 (1): 19–31.

Loftus, Elizabeth F, and Jacqueline E. Pickrell. 1995. “The Formation of False Memories.” *Psychiatric Annals* 25 (12): 720–25. https://doi.org/10.3928/0048-5713-19951201-07.

Mahr, Johannes B., and Gergely Csibra. 2018. “Why Do We Remember? The Communicative Function of Episodic Memory.” *Behavioral and Brain Sciences* 41. https://doi.org/10.1017/S0140525X17000012.

Malyon, Alan K. 1982. “Psychotherapeutic Implications of Internalized Homophobia in Gay Men: Journal of Homosexuality: Vol 7, No 2-3.” *Journal of Homosexuality* 7 (2–3): 59–69.

Martin, C. B., and Max Deutscher. 1966. “Remembering.” *The Philosophical Review* 75 (2): 161–96. https://doi.org/10.2307/2183082.

McDowell, James, and AE911Truth Staff. 2015. “60 Structural Engineers Cite Evidence for Controlled Demolition.” Architects & Engineers for 9/11 Truth. 2015. https://www.ae911truth.org/evidence/technical-articles/articles-by-ae911truth/199-60-structural-engineers.

McKinnon, Rachel. 2016. “Epistemic Injustice.” *Philosophy Compass* 11 (8): 437–46. https://doi.org/10.1111/phc3.12336.

Medina, José. 2011. “The Relevance of Credibility Excess in a Proportional View of Epistemic Injustice: Differential Epistemic Authority and the Social Imaginary.” *Social Epistemology* 25 (1): 15–35. https://doi.org/10.1080/02691728.2010.534568.

———. 2013. *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and Resistant Imaginations*. Studies in Feminist Philosophy. Oxford: Oxford University Press.

Medvecky, Fabien. 2018. “Fairness in Knowing: Science Communication and Epistemic Justice.” *Science and Engineering Ethics* 24 (5): 1393–1408. https://doi.org/10.1007/s11948-017-9977-0.

Meltzoff, Andrew N. 2005. “Imitation and Other Minds: The ‘like Me’ Hypothesis.” In *Perspectives on Imitation: Imitation, Human Development, and Culture*, edited by Susan Hurley and Nick Chater, m, 2:55–77. Cambridge, Mass.: MIT Press. https://mitpress.mit.edu/books/perspectives-imitation-volume-2.

Meyer, Ilan H. 2003. “Prejudice, Social Stress, and Mental Health in Lesbian, Gay, and Bisexual Populations: Conceptual Issues and Research Evidence.” *Psychological Bulletin* 129 (5): 674–97. https://doi.org/10.1037/0033-2909.129.5.674.

Meyer, Ilan H., and Laura Dean. 1998. “Internalized Homophobia, Intimacy, and Sexual Behavior among Gay and Bisexual Men.” In *Stigma and Sexual Orientation: Understanding Prejudice against Lesbians, Gay Men, and Bisexuals*, 160–86. Thousand Oaks, CA: SAGE Publications, Inc. https://doi.org/10.4135/9781452243818.n8.

Michaelian, Kourken. 2011a. “Is Memory a Natural Kind?” *Memory Studies* 4 (2): 170–89. https://doi.org/10.1177/1750698010374287.

———. 2011b. “The Epistemology of Forgetting.” *Erkenntnis* 74 (3): 399–424. https://doi.org/10.1007/s10670-010-9232-4.

———. 2011c. “Generative Memory.” *Philosophical Psychology* 24 (3): 323–42. https://doi.org/10.1080/09515089.2011.559623.

———. 2012a. “Metacognition and Endorsement.” *Mind & Language* 27 (3): 284–307. https://doi.org/10.1111/j.1468-0017.2012.01445.x.

———. 2012b. “(Social) Metacognition and (Self-)Trust.” *Review of Philosophy and Psychology* 3 (4): 481–514. https://doi.org/10.1007/s13164-012-0099-y.

———. 2016a. *Mental Time Travel: Episodic Memory and Our Knowledge of the Personal Past*. Cambridge, MA: MIT Press.

———. 2016b. “Confabulating, Misremembering, Relearning: The Simulation Theory of Memory and Unsuccessful Remembering.” *Frontiers in Psychology* 7 (November). https://doi.org/10.3389/fpsyg.2016.01857.

Michaelian, Kourken, and John Sutton. 2017. “Collective Memory.” In *Routledge Handbook of Collective Intentionality*, edited by Kirk Ludwig and Marija Jankovic, 140–51. London: Routledge.

Mills, Charles W. 1997. *The Racial Contract*. Ithaca, NY: Cornell University Press.

Minghella, Anthony. 1999. *The Talented Mr. Ripley*. Paramount Pictures.

Moshman, David. 2018. “Metacognitive Theories Revisited.” *Educational Psychology Review* 30 (2): 599–606. https://doi.org/10.1007/s10648-017-9413-7.

Nelson, Thomas O., and Louis Narens. 1994. “Why Investigate Metacognition?” In *Metacognition: Knowing about Knowing*, edited by Janet Metcalfe and Arthur P. Shimamura, 1–25. Cambridge: MIT Press.

Nichols, Shaun, and Stephen P. Stich. 2003. *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*. Oxford: Clarendon Press.
Nichols, Thomas M. 2017. *The Death of Expertise: The Campaign against Established Knowledge and Why It Matters*. New York: Oxford University Press.
Nisbett, Richard E., and Timothy DeCamp Wilson. 1977. "Telling More Than We Can Know: Verbal Reports on Mental Processes." *Psychological Review* 84 (3): 231–59.
Noë, Alva. 2005. "Against Intellectualism." *Analysis* 65 (4): 278–90.
O'Shea, Patrick. 2013. "Schools Locked down over 'Ambiguous Message' Taken as Threat." The Times. February 28, 2013. https://www.timesonline.com/article/20130228/News/302289698.
Parsons, J. 2012. "Cognitivism about Imperatives." *Analysis* 72 (1): 49–54. https://doi.org/10.1093/analys/anr132.
Peet, Andrew. forthcoming. "Testimonial Knowledge-How." *Erkenntnis*. https://doi.org/10.1007/s10670-018-9986-7.
———. 2017. "Epistemic Injustice in Utterance Interpretation." *Synthese* 194 (9): 3421–43. https://doi.org/10.1007/s11229-015-0942-7.
Pohlhaus, Gaile. 2012. "Relational Knowing and Epistemic Injustice: Toward a Theory of *Willful Hermeneutical Ignorance*." *Hypatia* 27 (4): 715–35. https://doi.org/10.1111/j.1527-2001.2011.01222.x.
Poston, Ted. 2009. "Know How to Be Gettiered?" *Philosophy and Phenomenological Research* 79 (3): 743–47. https://doi.org/10.1111/j.1933-1592.2009.00301.x.
———. 2016. "Know How to Transmit Knowledge?: Know How to Transmit Knowledge?" *Noûs* 50 (4): 865–78. https://doi.org/10.1111/nous.12125.
Proust, Joëlle, and Martin Fortier. 2018. *Metacognitive Diversity: An Interdisciplinary Approach*. Oxford University Press.
Ravenscroft, Ian. 2016. "Folk Psychology as a Theory." In *Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. https://plato.stanford.edu/entries/folkpsych-theory/.
Recanati, François. 2001. "What Is Said." *Synthese* 128 (1/2): 75–91.
Rehg, William, and Kent Staley. 2008. "The CDF Collaboration and Argumentation Theory: The Role of Process in Objective Knowledge." *Perspectives on Science* 16 (1): 1–25. https://doi.org/10.1162/posc.2008.16.1.1.
Riley, Evan. 2017. "What Skill Is Not." *Analysis* 77 (2): 344–54. https://doi.org/10.1093/analys/anx059.
Robins, Sarah. 2016a. "Misremembering." *Philosophical Psychology* 29 (3): 432–47. https://doi.org/10.1080/09515089.2015.1113245.
———. 2016b. "Representing the Past: Memory Traces and the Causal Theory of Memory." *Philosophical Studies* 173 (11): 2993–3013. https://doi.org/10.1007/s11098-016-0647-x.
———. 2017. "Confabulation and Constructive Memory." *Synthese*, February. https://doi.org/10.1007/s11229-017-1315-1.

Roediger, Henry L. 1980. "Memory Metaphors in Cognitive Psychology." *Memory & Cognition* 8 (3): 231–46. https://doi.org/10.3758/BF03197611.
Roediger, Henry L., and Kathleen B. McDermott. 1995. "Creating False Memories: Remembering Words Not Presented in Lists." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21 (4): 803–14.
Roland, Jane. 1958. "On 'Knowing How' and 'Knowing That.'" *The Philosophical Review* 67 (3): 379–88.
Rosenbaum, R. S., D. T. Stuss, B. Levine, and E. Tulving. 2007. "Theory of Mind Is Independent of Episodic Memory." *Science* 318 (5854): 1257–1257. https://doi.org/10.1126/science.1148763.
Rossiter, Margaret W. 1993. "The Matthew Matilda Effect in Science." *Social Studies of Science* 23 (2): 325–41.
Ryle, Gilbert. 1949. *The Concept of Mind*. London ; New York: Routledge.
Sant'Anna, André. 2018a. "Perception and Memory: Beyond Representationalism and Relationalism." Thesis, University of Otago. https://ourarchive.otago.ac.nz/handle/10523/8472.
———. 2018b. "Episodic Memory as a Propositional Attitude: A Critical Perspective." *Frontiers in Psychology* 9 (July). https://doi.org/10.3389/fpsyg.2018.01220.
Saul, Jennifer. 2013. "Implicit Bias, Stereotype Threat, and Women in Philosophy." In *Women in Philosophy*, edited by Katrina Hutchison and Fiona Jenkins, 39–60. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199325603.003.0003.
Schacter, Daniel L. 1999. "The Seven Sins of Memory: Insights From Psychology and Cognitive Neuroscience." *American Psychologist* 54: 182–203.
Schacter, Daniel L., and Donna Rose Addis. 2007. "The Cognitive Neuroscience of Constructive Memory: Remembering the Past and Imagining the Future." *Philosophical Transactions: Biological Sciences* 362 (1481): 773–86.
Schacter, Daniel L., Kenneth A. Norman, and Wilma Koutstaal. 1998. "The Cognitive Neuroscience of Constructive Memory." *Annual Review of Psychology* 49 (1): 289–318.
Schnider, Armin. 2018. *The Confabulating Mind: How the Brain Creates Reality*. Oxford: Oxford University Press.
Schraw, Gregory, and Rayne Sperling Dennison. 1994. "Assessing Metacognitive Awareness." *Contemporary Educational Psychology* 19 (4): 460–75. https://doi.org/10.1006/ceps.1994.1033.
Schraw, Gregory, and David Moshman. 1995. "Metacognitive Theories." *Educational Psychology Review* 7 (4): 351–71. https://doi.org/10.1007/BF02212307.
Scott-Phillips, Thom. 2015. *Speaking Our Minds*. Palgrave.
Sellars, Wilfrid. 1955. *Empiricism and the Philosophy of Mind*. Cambridge, Mass.: Cambridge University Press.
Senor, Thomas D. 2007. "Preserving Preservationism: A Reply to Lackey." *Philosophy and Phenomenological Research* 74 (1): 199–208.

Shannon, Claude Elwood, and Warren Weaver. 1949. *The Mathematical Theory of Communicaton*. University of Illinois Press.

Shanton, Karen. 2011. "Memory, Knowledge and Epistemic Competence." *Review of Philosophy and Psychology* 2 (1): 89–104. https://doi.org/10.1007/s13164-010-0038-8.

Shanton, Karen, and Alvin Goldman. 2010. "Simulation Theory." *Wiley Interdisciplinary Reviews: Cognitive Science*, February, n/a-n/a. https://doi.org/10.1002/wcs.33.

Shaw, Julia, and Stephen Porter. 2015. "Constructing Rich False Memories of Committing Crime." *Psychological Science* 26 (3): 291–301. https://doi.org/10.1177/0956797614562862.

Sherman, Benjamin R. 2016. "There's No (Testimonial) Justice: Why Pursuit of a Virtue Is Not the Solution to Epistemic Injustice." *Social Epistemology* 30 (3): 229–50. https://doi.org/10.1080/02691728.2015.1031852.

Sigurdsson, Jon F., and Gisli H. Gudjonsson. 1996. "The Psychological Characteristics of 'False Confessors'. A Study among Icelandic Prison Inmates and Juvenile Offenders." *Personality and Individual Differences* 20 (3): 321–29. https://doi.org/10.1016/0191-8869(95)00184-0.

Smith, Dorothy E. 1974. "Women's Perspective as a Radical Critique of Sociology." *Sociological Inquiry* 44 (1): 7–13. https://doi.org/10.1111/j.1475-682X.1974.tb00718.x.

———. 1987. *The Everyday World As Problematic: A Feminist Sociology*. Boston: Northeastern University Press.

Snowdon, Paul. 2003. "Knowing How and Knowing That: A Distinction Reconsidered," 29.

Sosa, Ernest. 1991. *Knowledge in Perspective: Selected Essays in Epistemology*. Cambridge University Press.

Spear, Andrew D. 2020. "Gaslighting, Confabulation, and Epistemic Innocence." *Topoi* 39: 229–41. https://doi.org/10.1007/s11245-018-9611-z.

Speight, Suzette L. 2007. "Internalized Racism: One More Piece of the Puzzle." *The Counseling Psychologist* 35 (1): 126–34. https://doi.org/10.1177/0011000006295119.

Sperber, Dan, and Deirdre Wilson. 1986. *Relevance: Communication and Cognition*. Oxford: Basil Blackwell Ltd.

Spreng, R. Nathan, Raymond A. Mar, and Alice S.N. Kim. 2008. "The Common Neural Basis of Autobiographical Memory, Prospection, Navigation, Theory of Mind, and the Default Mode: A Quantitative Meta-Analysis." *Journal of Cognitive Neuroscience* 21 (3): 489–510.

Squire, Larry R. 1992. "Declarative and Nondeclarative Memory: Multiple Brain Systems Supporting Learning and Memory." *Journal of Cognitive Neuroscience* 4 (3): 232–43. https://doi.org/10.1162/jocn.1992.4.3.232.

———. 2004. “Memory Systems of the Brain: A Brief History and Current Perspective.” *Neurobiology of Learning and Memory* 82 (3): 171–77. https://doi.org/10.1016/j.nlm.2004.06.005.

Squire, Larry R., and Stuart Zola-Morgan. 1988. “Memory: Brain Systems and Behaviour.” *Trends in Neurosciences* 11 (4): 170–75. https://doi.org/10.1016/0166-2236(88)90144-0.

Staley, Kent W. 2007. “Evidential Collaborations: Epistemic and Pragmatic Considerations in ‘Group Belief.’” *Social Epistemology* 21 (3): 321–35. https://doi.org/10.1080/02691720701674247.

Stanley, Jason. 2011. *Know How*. Oxford: Oxford University Press.

Stanley, Jason, and Timothy Willlamson. 2001. “Knowing How.” *Journal of Philosophy* 98 (8): 411–44. https://doi.org/10.2307/2678403.

Steele, Claude M. 1997. “How Stereotypes Shape Intellectual Identity and Performance.” *American Psychologist* 52 (6): 613–29.

Steele, Claude M., and Joshua A. Aronson. 1995. “Stereotype Threat and the Intellectual Test Performance of African Americans.” *Journal of Personality and Social Psychology* 69 (5): 797–811.

———. 2004. “Stereotype Threat Does Not Live by Steele and Aronson (1995) Alone.” *American Psychologist* 59 (1): 47–55.

Steele, Claude M., Steven J. Spencer, and Joshua Aronson. 2002. “Contending with Group Image: The Psychology of Stereotype and Social Identity Threat.” In *Advances in Experimental Social Psychology*, 34:379–440. Academic Press. https://doi.org/10.1016/S0065-2601(02)80009-0.

Stickgold, Robert. 2013. “Parsing the Role of Sleep in Memory Processing.” *Current Opinion in Neurobiology* 23 (5): 847–53. https://doi.org/10.1016/j.conb.2013.04.002.

Strauss, Valerie. 2013. “Schools Put on Lockdown When ‘Fresh Prince’ Song Is Misinterpreted.” Washington Post. March 3, 2013. https://www.washingtonpost.com/news/answer-sheet/wp/2013/03/03/misheard-fresh-prince-song-sparks-schools-lockdown/.

Sutton, John. 1998. *Philosophy and Memory Traces: Descartes to Connectionism*. Cambridge [England] ; New York: Cambridge University Press.

Sutton, John, and Carl Windhorst. 2009. “Extended and Constructive Remembering: Two Notes on Martin and Deutscher.” *Crossroads* 4 (1): 79–91.

Szymanski, Dawn M., Arpana Gupta, Erika R. Carr, and Destin Stewart. 2009. “Internalized Misogyny as a Moderator of the Link between Sexist Events and Women’s Psychological Distress.” *Sex Roles* 61 (1–2): 101–9. https://doi.org/10.1007/s11199-009-9611-y.

Tait, Amelia. 2016. “The Movie That Doesn’t Exist and the Redditors Who Think It Does.” New Statesman. December 21, 2016. https://www.newstatesman.com/science-tech/internet/2016/12/movie-doesn-t-exist-and-redditors-who-think-it-does.

Talland, George. 1961. "Confabulation in the Wernicke-Korsakoff Syndrome." *Journal of Nervous and Mental Disease* 132: 361–81. https://doi.org/10.1097/00005053-196105000-00001.
———. 1965. *Deranged Memory: A Psychonomic Study of the Amnesic Syndrome*. New York: Academic Press.
Tanesini, Alessandra. 2018. "Collective Amnesia and Epistemic Injustice." In *Socially Extended Epistemology*, edited by J. Adam Carter, Andy Clark, Jesper Kallestrup, S. Orestis Palermos, and Duncan Pritchard, 1:195–219. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198801764.003.0011.
Theiner, Georg. 2013. "Transactive Memory Systems: A Mechanistic Analysis of Emergent Group Memory." *Review of Philosophy and Psychology* 4 (1): 65–89. https://doi.org/10.1007/s13164-012-0128-x.
Tollefsen, Deborah Perron. 2015. *Groups as Agents*. Cambridge: Polity Press.
Tulving, Endel. 2007. "Are There 256 Different Kinds of Memory?" In *The Foundations of Remembering: Essays in Honour of Henry L. Roediger, III*, edited by J.S. Nairne, 39–52. New York: Psychology Press.
Tuomela, Raimo. 1992. "Group Beliefs." *Synthese* 91 (3): 285–318.
Turnbull, William. 2003. *Language in Action: Psychological Models of Conversation*. Psychology Press.
u/AscendedMinds. 2017. "Scientists Believe Parallel Universes ARE Interacting. Is This the Cause of the 'Mandela Effect'?" *Reddit r/MandelaEffect*. https://www.reddit.com/r/MandelaEffect/comments/6eq5k7/scientists_believe_parallel_universes_are/.
u/AutoModerator. 2018. "Did You Discover a Possible New Mandela Effect? Post It Here! (Weekly Discussion) (2018-06-24)." *Reddit r/Mandela Effect*. https://www.reddit.com/r/MandelaEffect/comments/8tfyqc/did_you_discover_a_possible_new_mandela_effect/.
u/Denominax. 2017. "Mandela Effect Wiki." Reddit r/MandelaEffect. May 18, 2017. https://old.reddit.com/r/MandelaEffect/wiki/index.
u/DonnaGail. 2017. "Shazam / Shazaam with Sinbad Was Real and Here Is All the Movie Information." *Reddit r/Shazaam*. https://www.reddit.com/r/Shazaam/comments/5m8o02/shazam_shazaam_with_sinbad_was_real_and_here_is/ddjzs8p.
u/EpicJourneyMan. 2016. "The Sinbad Genie Movie - Complete Analysis." *Reddit r/Mandela Effect*. https://www.reddit.com/r/MandelaEffect/comments/55f5rt/the_sinbad_genie_movie_complete_analysis/.
———. 2018. "All Of My Coworkers Remember Shazam." *Reddit r/Mandela Effect*. https://www.reddit.com/r/MandelaEffect/comments/7nhdjt/all_of_my_coworkers_remember_shazam/ds73ebi.
u/ezydown. 2017. "254 Confirmed Mandela Effects: List." *Reddit r/Mandela Effect*. https://www.reddit.com/r/MandelaEffect/comments/6p6zwd/254_confirmed_mandela_effects_list/dknskwm/.

u/Fae_Leaf. 2018. "All Of My Coworkers Remember Shazam." *Reddit r/Mandela Effect*. https://www.reddit.com/r/MandelaEffect/comments/7nhdjt/all_of_my_coworkers_remember_shazam/.

u/Jhoobie. 2017. "What Is the Mandela Effect?" *Reddit r/OutOfTheLoop*. https://www.reddit.com/r/OutOfTheLoop/comments/5m644b/what_is_the_mandela_effect/dc1ib3d.

u/manafrmhvn. 2018. "All Of My Coworkers Remember Shazam." *Reddit r/Mandela Effect*. https://www.reddit.com/r/MandelaEffect/comments/7nhdjt/all_of_my_coworkers_remember_shazam/ds6josi.

u/melossinglets. 2017. "254 Confirmed Mandela Effects: List." *Reddit r/Mandela Effect*. https://www.reddit.com/r/MandelaEffect/comments/6p6zwd/254_confirmed_mandela_effects_list/dko45es.

u/shazaamthemovie. 2017. "Shazam / Shazaam with Sinbad Was Real and Here Is All the Movie Information." *Reddit r/Shazaam*. https://www.reddit.com/r/Shazaam/comments/5m8o02/shazam_shazaam_with_sinbad_was_real_and_here_is/.

u/squidink20. 2018. "Shazaam Was Probably Buried/Erased from History by Executives." *Reddit r/MandelaEffect*. https://www.reddit.com/r/MandelaEffect/comments/92y4g2/shazaam_was_probably_buriederased_from_history_by/.

u/ThadeusOfNazareth. 2016. "What Is Going on with Mother Teresa?" *Reddit r/MandelaEffect*. https://www.reddit.com/r/MandelaEffect/comments/516nen/what_is_going_on_with_mother_teresa/.

u/TimmehTheShpee. 2018. "Monopoly Man, I Swear to God He Has a Monocle." *Reddit r/MandelaEffect*. https://www.reddit.com/r/MandelaEffect/comments/9mhqny/monopoly_man_i_swear_to_god_he_has_a_monocle/.

Vendler, Zeno. 1972. *Res Cogitans*. Ithaca, NY: Cornell University Press.

Vicente, Agustín, and Fernando Martínez-Manrique. 2008. "Thought, Language, and the Argument From Explicitness." *Metaphilosophy* 39 (3): 381–401. https://doi.org/10.1111/j.1467-9973.2008.00545.x.

Wegner, Daniel M. 2002. *The Illusion of Conscious Will*. Cambridge, Mass: MIT Press.

Werning, Markus, and Sen Cheng. 2017. "Taxonomy and Unity of Memory." In *The Routledge Handbook of Philosophy of Memory*, edited by Sven Bernecker and Kourken Michaelian, 7–20. London: Routledge.

Williams, John N. 2008. "Propositional Knowledge and Know-How." *Synthese* 165 (1): 107–25. https://doi.org/10.1007/s11229-007-9242-1.

Wilson, Timothy D. 2002. *Strangers to Ourselves: Discovering the Adaptive Unconscious*. Cambridge, Mass: Harvard University Press.

Wimsatt, William C. 1986. “Forms of Aggregativity.” In *Human Nature and Natural Knowledge: Essays Presented to Marjorie Grene on the Occasion of Her Seventy-Fifth Birthday*, edited by Alan Donagan, Anthony N. Perovich, and Michael V. Wedin, 259–91. Boston Studies in the Philosophy of Science. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-009-5349-9_14.

Wray, K. Brad. 2001. “Collective Belief and Acceptance.” *Synthese* 129 (3): 319–33.

Wright, Sarah. forthcoming. “Epistemic Harm and Virtues of Self-Evaluation.” *Synthese*. https://doi.org/10.1007/s11229-018-01993-x.

Zagzebski, Linda Trinkaus. 2012. *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199936472.001.0001.