Remembering in Communication

A novel account of the architecture and function of human episodic memory

Johannes B. Mahr

Submitted to:

Central European University

Department of Cognitive Science

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Cognitive Science

Primary supervisor: Gergely Csibra

Secondary supervisor: Dan Sperber

Budapest, Hungary

2019

ABSTRACT

Remembering is a fundamental component of human life and cognition. Humans spend a sizable portion of their lives with thinking and talking about their past experiences. The past and our memories of it seem to be particularly important to us. In this thesis, I develop an account of what 'remembering' is, why we think about remembering the way we do, why humans remember at all, and why it plays such an important role in human culture and social life.

In Chapter 1, I introduce the intuitive notion of remembering which has dominated philosophical discourse in the last few decades. I then move on to explain how the intuitions underlying this notion are cognitively produced by focusing on the mechanisms and evolution of human episodic memory.

First, in Chapter 2, I give an account of the cognitive architecture of episodic simulation – the cognitive system producing the contents of episodic memory. I argue that episodic memory is just one specific output of the wider ability of episodic simulation (i.e. the capacity to produce mental imagery about events outside our sensory scope).

Second, in Chapter 3, I focus on episodic memory specifically. Episodic memory goes beyond the outputs of episodic simulation because it includes a representation of its own causal history. When we represent an event in episodic memory, we do not just represent the event itself but also how we came to know about it, namely, through our own experience.

Third, in Chapter 4, I give an account of the evolved function of the episodic memory system. That is, I explain why episodic memory has a metarepresentational structure including information about its origin in first-hand experience. I argue that this metarepresentational structure functions to allow us to determine when we can lay claim to epistemic authority about the past in communication.

Finally, in Chapter 5, I ask why this ability would have been useful in the representation of particular past events. That is, I aim to answer the question why the ability to determine whether one has epistemic authority is particularly important for representations of and claims about the past. Here, I argue that for humans the past is special because it is often the only way we can determine present social realities. For humans, certain events (like promises, transfers of ownership, etc.) are conceived of as causes of social entities like commitments, entitlements, obligations, and accountabilities. The representation of token causes for such specific social effects is crucial because they commonly do not leave concomitant, traceable changes in the physical environment. Social effects like commitments, entitlements and obligations often consist only in mental representations and depend on interpersonal agreement to be effective. The only way the existence of such social effects can be negotiated is by recourse to their specific cause in the past.

This does not only explain why history has such high importance to us as individuals and members of social groups, but also why humans have culturally developed a large range of technologies for making specific events with consequential social effects public, documented, and traceable, as well as why claims about the past are a continuous source of conflict.

ACKNOWLEDGMENTS

First, I would like to thank my supervisor Gergely Csibra. This thesis was built slowly over the past five years over a course of frustrating failures and dead ends. The only reason it became what it is now is due to Gergo's open-mindedness, patience, ingenuity and unwavering support. Whenever I felt a lack of intellectual inspiration or could not find the answer to a crucial question, I could count on his doors being open. Most of the time, I would leave his office with a head full of new exciting ideas. Those many conversations we had over the past years made me into the academic I am today and shaped my view on many fundamental issues.

Second, I want to thank my second supervisor Dan Sperber, who, through his advice in person and in writing, helped me navigate some of the central questions of this project.

Next, of course, I am grateful to my family. Without my parents, Albrecht and Brigitta, who always supported me in every way they could and never stopped believing in me, I would have never been able to complete this thesis. I also want to thank my sister, Ines, whose friendship and loving concern have been extremely dear to me.

Further, I want to express my gratitude to two other intellectual mentors whose work, advice and support has helped me in many ways over the last years: Hugo Mercier and Kourken Michaelian. I am grateful for the many occasions on which I was able to seek advice, or discuss ideas with each of them. Their intellectual influence can be felt throughout this thesis.

Finally, I want to thank the many friends without whom I would have never been able to endure the PhD-lifestyle. Maria, who has always been there and with whom I was able to share so many adventures, pains and joys. Johannes, without whom my time in Budapest

would have lacked a lot of color and who continues to be a role model for me for his integrity and compassion. Clara, whose personal and intellectual autonomy continues to be a source of inspiration. Jakob, whose long friendship has given continuity to my life despite otherwise changing landscapes. Basia, who always encouraged me and was excited to offer advice or to discuss scientific ideas. All the members of MigSzol who I was able to learn from so much and share so many exciting, hard and valuable lessons with. The other members of our flatshare in Berlin, Kascha, Charleen, and Marvin, who have made my life here beautiful and the process of writing this thesis so pleasurable in the last few months (Achtsamkeit und Dankbarkeit). And the many other friends who have been part of my life during the course of my PhD: my life would be emptier without you.

DECLARATION OF AUTHORSHIP

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials written by another person, or which have been accepted for the award of any other degree or diploma at Central European University or any other educational institution, except where due acknowledgment is made in the form of bibliographical reference.

_____

Johannes B. Mahr

TABLE OF CONTENTS

Versions of parts of this thesis have been published, submitted for publication, or are currently under review for publication: Chapter 2 is currently under review for publication in *Psychological Review*. Chapters 3 and 4 have been published as Mahr, J.B. & Csibra, G. (2018). Why do we remember? The communicative function of episodic memory. *Behavioral and Brain Sciences*, *41*. Finally, Chapter 5 has been submitted for publication to *Perspectives on Psychological Science.*

CHAPTER 1

GENERAL INTRODUCTION: WHAT IS REMEMBERING?

This thesis is about remembering. It seems fitting then to start by asking the question of what remembering is. While part of the aim of this thesis is to develop an account that answers this question, it is worth spending some time first to clarify what the question actually amounts to. Both in psychology and in philosophy, thought about remembering has a long history, albeit the questions that have been asked about it have been (maybe unsurprisingly) markedly different.

In philosophy, 'remembering' is commonly simply taken to be equivocal to the exercise of memory itself. And the questions that have most prominently exercised philosophers of memory have been ontological and epistemological ones: What is it that we should call a memory? Does memory produce knowledge? What distinguishes memory from other forms of knowing? In this vein of thought, one particular paper by Martin & Deutscher (1966) has been particularly influential. In it, the authors develop an account of the nature of memory (i.e. what kind of things we should call memories) known as the 'causal theory of memory'. The causal theory is named thus because it makes memory contingent on the existence of an appropriate causal link between a past event and a current representation of this event. For the last few decades, the causal theory has been the dominant view about memory in philosophy. Only recently, and spurred by evidence from psychology, it has come under fire by theorists questioning whether it is compatible with the psychological mechanisms of remembering (Michaelian & Robins, 2018; Robins, 2016). In section 2 I will shortly introduce

the causal theory and specify in what way the ideas that have been proposed in the philosophy of memory are relevant to the project of this thesis. My interest in this thesis, however, will decidedly not be to get involved in the kinds of debates that have been going on in the metaphysics of memory.

Rather, I will take the intuitions that have given rise to the conditions philosophers have commonly applied for something to count as an instance of remembering to be rooted in the way one specific psychological system operates: Episodic memory. That is, I am interested in a psychological theory of remembering. Thus, I will take the view that remembering is a psychological category: Remembering as an object of psychological investigation is any mental state that presents itself as being 'remembered'. Of course, it is all but clear what remembering should be from a psychological perspective even disregarding the epistemological concerns philosophers have been worrying about (to the extent that this is possible). In section 1, I will therefore shortly review the psychological view of human memory as encapsulated in the classical taxonomy of memory systems. I will specify episodic memory as the target of investigation in this thesis while staying agnostic about whether this cognitive system deserves to be called 'memory' in an ontological sense. In section 3, I will then give an outlook and a short overview over what I aim to achieve in the remainder of this thesis.

One more thing is worth noting before I continue: this thesis is, an exercise in 'theoretical psychology', that is, of theory development within cognitive science. Therefore, while I will draw on debates, concepts and methods from philosophy in many places, I do not view this thesis as being primarily aimed at answering philosophical questions. Rather, it is my aim to contribute to psychology, that is, to our naturalistic understanding of how the mind works. Such theory building can be important for its own sake, allowing us to make sense of an

otherwise disconnected space of empirical results. However, any psychological theory is only worth its salt to the extent that it can formulate predictions which can inform empirical research. Following Dennett's (2009) advice, it will therefore be my aim to both, build on empirical results in my theory building and specify empirical predictions of those theories wherever possible.

1. THE PSYCHOLOGY OF MEMORY

In the last few decades, most psychological research on memory has been structured by the idea that human memory is not a unitary capacity but can be separated into different systems (Squire & Zola-Morgan, 1988, Squire, 1992; Squire, 2004). A '*system*', on this view, is a functionally and structurally integrated computational module[1] which can be distinguished on the basis of its mechanisms, the contents it processes, and its neural implementation from other such modules. In other words, a cognitive system, can be described and distinguished from other systems in terms of (1) the information processing tasks it performs, (2) the computations it uses to perform this task, and (3) the way these computations are implemented in the brain (Michaelian, 2011a; 2016; Marr, 1982).

While it seems fairly straight-forward to identify a cognitive *system*, it is less clear what makes a system a *memory* system. Commonly, cognitive systems have been thought of as 'memorial' whenever they implement the encoding, storage and retrieval of information. In other words, a system is a memory system if the contents it uses originate in an initial act of learning, are retained in one form or another and eventually retrieved (Klein, 2015).

[1] I am using the term 'module' here without wanting to commit to any of the theoretical baggage that is often associated with it (e.g. Barrett & Kurzban, 2006).

According to the classical taxonomy, memory systems are thought to be hierarchically organized such that each system instantiates its own 'kind' of memory serving independent functions in learning and behavior. Therefore, each memory system, while normally working together with other systems, should be functionally and structurally independent of other systems. Evidence for this view came initially from behavioral studies with amnesic patients (Squire, 2004). For example, it has been shown that amnesic patient H.M. could learn novel motor skills (Milner, 1962), was subject to perceptual priming and could learn novel semantic facts (e.g. Tulving et al. 1991) without being able to recall anything about the situations in which such learning occurred. The fact that these different types of learning could be dissociated in this way was therefore taken as evidence for the idea that the ability to recall events (or 'episodic memory') must be functionally and structurally distinct from other types of memory (Tulving, 1983). Similar kinds of evidence have been produced for the other proposed distinctions in the taxonomy that makes up the memory systems view. Figure 1 displays these memory systems according to the classical taxonomy proposed by Squire (2004).

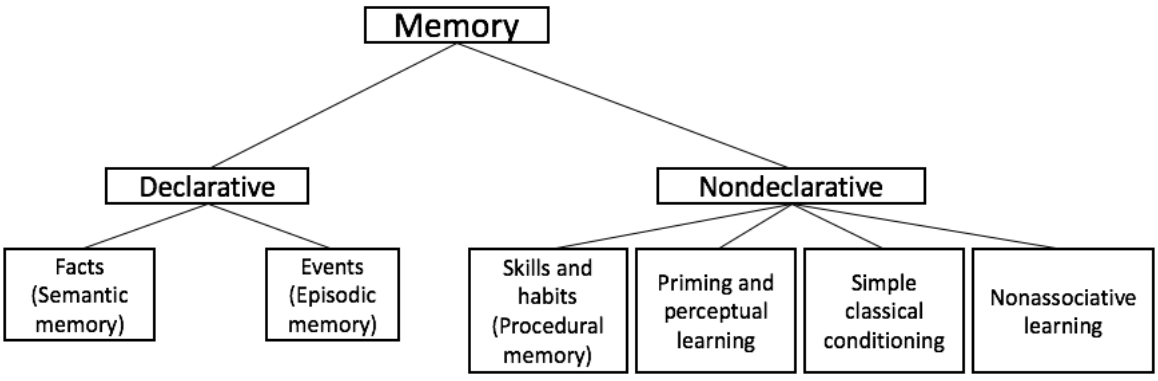

Figure 1: The classical taxonomy of memory systems as proposed by Squire (2004).

The first, basic, distinction that makes up this taxonomy is that between declarative and nondeclarative types of memory. As the name suggests, contents in declarative systems can

be ‘declared’ (and are therefore thought to be explicit and conscious) whereas those in nondeclarative systems cannot (and are therefore thought to be implicit and unconscious). Declarative and non-declarative systems are then further divided according to what kind of information is processed. On the nondeclarative side, the learning of skills and habits (procedural learning), perceptual relations (priming), stimulus-response associations (conditioning), and non-associative/reflexive learning are distinguished. On the declarative side, generic knowledge (semantic) is distinguished form memory for events (episodic).

In this thesis, I will be concerned with what Squire has here described as ‘memory for events’, that is, episodic memory. As will become clear, it is far from straight-forward how episodic memory and its distinction from semantic memory should be characterized. But I want to put these considerations on hold for now and instead consider the memory systems taxonomy in light of the ontology of memory it suggests.

Note that the idea of different ‘kinds’ of memory might be applied to either the ‘top-layer’ of the ‘bottom-layer’ of this taxonomy. That is, we might either think that the relevant distinction between memory systems is according to whether they are declarative or non-declarative or as to what type of information is processed. One might think that this question is solved by simply investigating at which level the hierarchy is functionally and structurally distinguishable. If we find, as seems to be the case, that the processing of each type of information described in the bottom-layer of the hierarchy is indeed functionally independent from the others, we should take this layer to represent actual distinctions between memory systems (e.g. Fernandez, 2018). Establishing that there are functionally and structurally integrated capacities at the bottom level of the hierarchy which can clearly be distinguished from each other is to show that we can apply the term ‘system’ to each. It is, however, not yet to show that we can apply the term ‘memory’ to them.

As mentioned above, psychologists have commonly assumed that, in order for a system to count as memory, it should implement the encoding, storage and retrieval of information. It does, however, not seem obvious in what way we can understand the operation of nondeclarative systems as 'encoding', 'storage' and 'retrieval' (Michaelian, 2011a). Procedural learning, for example, does not seem to rely on representational states at all (Schacter & Tulving, 1994). Indeed, if we look at what the systems grouped under the non-declarative heading actually do, it seems more appropriate to refer to them as *learning*, not memory, systems. As Klein (2015) has pointed out, however, even if we restrict the concept of memory in this way, much of our mental functioning will turn out to be memorial. The term 'memory' might then be bloated to an extent so as to become virtually vacuous. In fact, this point would not only apply to mental processes but to cultural ones, too (Berliner, 2005). A book might be said to encode, store and allow for the retrieval of a fact (say), but it seems odd to therefore call it 'memory' (even though Clark and Chalmers (1998) might disagree).

There are two ways in which this issue has been addressed: On the one hand, people have tried to specify under which conditions memory could be taken to be a natural kind and to what extent the cognitive processes underlying various systems conform to this specification (e.g. Cheng & Werning, 2016; 2017; Michaelian, 2011a; 2015). On the other hand, in philosophy there has been a long tradition of searching for necessary and sufficient conditions under which the term 'memory' can be applied to a mental state (Michaelian & Robins, 2018; Bernecker, 2008). Both approaches, the search for memory as a natural kind and the attempts to define memory in terms of necessary and sufficient condition, have arguably traded on some deep, a priori intuitions about the way remembering (i.e. the exercise of memory) operates and what it amounts to. Only recently have authors made the

attempt to reconcile these intuitions about remembering with evidence gathered from the cognitive sciences about how systems traditionally understood as memorial actually operate. I will not aim to contribute to this discourse. Instead, I will suggest that both the intuition that memory must be a natural kind as well as the idea that there are necessary and sufficient conditions specifying when we can call something a memory are results of the operation of one system: Episodic memory.

In this thesis, I will therefore offer an account of how episodic memory operates. That is, I will only be concerned with giving an account of the psychological processes underlying one of the systems in the classical psychological taxonomy. I will assume that episodic memory is, indeed, a cognitive system, that is functionally and structurally integrated. I will, however, make no assumptions or claims about whether the term 'memory' can be rightfully applied to it. Nonetheless, I will take the intuitions that lie at the foundation of the debate about the nature of remembering as an explanandum. And I will aim to show that the operation of episodic memory is at the core of what causes us to have the intuitions we do about what remembering is. In other words, episodic memory seems to give rise to the self-attribution of remembering and I will argue that this self-attribution gives rise to our intuitions about what remembering/memory is. One of the questions I will therefore aim to answer in what follows is: Why do we have the intuitions we do about what remembering is and how it operates?

Of course, in order to explain our memorial intuitions, we should first have an idea of what they consist in. I will take some central debates in the philosophy of memory to be an outcome of these intuitions. Therefore, these debates can serve as data about as to what our intuitions about memory are. In the next section, I will therefore shortly give an account of what I take to be the standard view of memory in contemporary philosophy. The

conditions which this standard view applies to the identification of memory can then serve as a target for explanation when we consider the representational structure of episodic memory. In particular, I take one intuition to be particularly central to our understanding of memory: The idea that in order for someone to remember something, their current mental state has to have been caused by their past experience (Michaelian, 2015).

2. THE PHILOSOPHY OF MEMORY

Memory, under any definition, is central to the human mind and, as such, has been a subject of philosophical investigation for thousands of years. Aristotle and Plato developed early theories about memory and it has been a central concern for many early modern philosophers such as Hume, Reid, and Locke. This, however, is not the place to review the entire history of the philosophy of memory. I want, instead, to shortly give an overview of one central debate in the contemporary philosophy of memory.

This debate is concerned with identifying the ontological *nature of memory*. When asking about the nature of memory, one way to understand this question would be as 'how can we identify something as a memory?', that is what are the conditions under which someone can justifiably be said to remember something?

2.1. THE INTUITIVE NOTION OF REMEMBERING

It is not always clear whether philosophers interested in memory want to give an account of the metaphysics of memory, that is, what memories actually are, independently of how our psychology produces them, or if they want to give an account of the psychology of memory,

that is, how our mind produces mental states that are memories. Traditionally, the philosophical and psychological investigations of remembering have proceeded mostly in parallel without much interaction. Only in recent years, the discussion about the nature of memory has started to revolve around the attempt to give an account of 'what remembering is' that is compatible with evidence from psychology while also allowing us to ontologically distinguish between imagining and remembering (for review see Werning & Cheng, 2017).

Thus, the interaction between philosophy and psychology has proceeded mostly in one direction in so far as philosophers have used insights from psychology to inform their own accounts. There is, however, also the possibility of interaction in the other direction: The way philosophers have thought about the ontology of memory can be of interest to psychologists. Philosophical thinking can clarify precisely what our commonsense understanding of a given phenomenon is by probing what our intuitions about a certain concept turn out to be. Such explications of commonsense can then become the subject of theorizing about the origin of these intuitions; that is, why commonsense is commonsense. Of course, there have been many different philosophical accounts of the nature of memory (for reviews see e.g. Bernecker, 2008; 2010; Michaelian & Robins, 2018). Nonetheless, in the last few decades, the causal theory proposed by Martin and Deutscher (1966) has arguably been the most widely accepted account of what remembering is. I will therefore take it that the causal theory is a somewhat accurate guide as to what our intuitions about the nature of remembering are. Moreover, proponents of the causal theory commonly take it to be a strength of their view that it aligns with our intuitions about memory. In fact, it is hard to see what other source the necessary and sufficient conditions which the causal theory applies to memory could have.

The causal theory trades on the notion that when I remember an event, I do so *because I experienced it*. In other words, according to the causal theory (as the name suggests) remembering comes down to the existence of a causal connection between one's original experience of an event and one's current representation of it. Without getting into the technical details of the argument, the causal theory states that someone remembers an event, iff

(1) the event occurred,

(2) she now represents the event,

(3) she originally experienced the event as it took place,

(4) her current representation of the event sufficiently resembles her original experience of it, and

(5) there is a causal connection between that original experience and her current representation of the event.

Condition (1) is meant to ensure factivity, that is, when I remember an event, I commonly believe that this event occurred. Condition (2) expresses that, memories must be mental states; that is, when I remember, I am mentally representing a past event. Further, I can only be said to remember an event if I experienced this event in the past (Condition 3) and am now representing it accurately, that is, in the way I originally experienced it (Condition 4). These conditions are meant to ensure the distinction between cases in which a subject only 'seems to remember', that is, entertains a mental state that presents itself as a memory but is not actually based on a past experience or 'remembers falsely', that is, remembers in a way that does sufficiently resemble the original experience. Finally, Condition 5 is meant to

enable the distinction between cases in which someone takes herself to imagine even though she is in fact remembering and cases in which one takes oneself to remember even though one is in fact imagining. According to the causal theory, the distinction between imagining and remembering in these cases is enabled by whether there was an appropriate casual connection between a given past experience and the present event representation under scrutiny. Remembering can only be said to occur if one's current representation was caused (in the right way) by one's past experience.

A large amount of effort has been spent on clarifying what kind of causal connection would be appropriate so as to ensure that Condition 5 could distinguish between the remembering and imagining of events (Robins, 2016). According to Martin and Deutscher, this causal connection should be viewed in terms of a 'memory trace' that is some stored analog of the original event. In fact, the history of the philosophy of memory is full of analogies (wax tablets, file drawers etc.) trying to explain the operations of memory in terms of an analog of original experience stored across time in the mind. Since debates about what the appropriate kind of causal connection should look like are intended to allow for an ontological distinction between remembering and imagining, they are not relevant to the project I am interested in here. What I want to emphasize here is that the intuition underlying the debate about the nature and existence of memory traces in the first place is that *one's current representation of the event must have been caused by one's past experience of it*. It is this intuition that I take to be central to our intuitive notion of remembering.

3. Outlook

Having spelled out the intuitive notion of remembering, it can give us an idea of the kind of thing I will be trying to explain in a naturalistic manner. We are looking for a cognitive system that outputs a representation which could ground all of the intuitions encapsulated in the causal theory described above. That is, these representations should be of a past event, which actually occurred, acquired through one's own past experience and which allows one to relive this event in the present to some extent.

In chapters 2 and 3 I will be concerned with giving an account of the cognitive architecture producing these kinds of representations. Since I take episodic memory to be the cognitive system underlying remembering, this account will be of the cognitive architecture of episodic memory.

Chapter 2 will be concerned with the cognitive system that produces the content of episodic memory, that is event representations. This 'episodic simulation' system can be understood as a mechanism that allows for the simulation of event representations including the temporal and spatial relations in a quasi-perceptual manner. These representations can differ in a number of ways: We can simulate actual events, counterfactual events, future events, past events and so on. Part of the goal of chapter 2 will be to give an account of how the episodic simulation system can produce such a wide range of event representations.

One of the kinds of event representations which episodic simulation can produce are representations of specific, past, actual events. That is, events that one believes to have actually occurred, in the past at one specific point in time. It is these kinds of event representations, which, I claim, constitute the content of episodic memory. However, episodic memory is not simply a mental simulation of a specific, past, actual event. In

chapter 3, I will therefore zoom in more closely on how episodic memory results from and goes beyond the operations of episodic simulation. I will argue that episodic memory goes beyond a simple output of episodic simulation because it includes source information: When we remember an event episodically, we do not just believe that it occurred, we also believe that our current representation of this event was caused by our past, first-hand experience of it.

Having established the cognitive architecture underlying episodic memory, we can then ask, why do human adults possess such a system? That is, why do we represent past events in a way that includes information about how we came to know about them (namely, through having experienced them)? Chapter 4 will address this question by arguing for a novel account of the evolved function of episodic memory. I will argue that the kind of source information episodic memory includes (that is, about information having been acquired through first-hand experience) is crucial in allowing us to transmit information about past events to others and evaluate such information transmitted by others in communication. First-hand experience bestows epistemic authority and it is often only by such authority that we can decide what to believe about the past.

Why, however, should it be important to decide what to believe about the past? Why should the past be so important for human beings that we have developed dedicated cognitive mechanism (episodic memory) allowing us to gauge the extent to which we have epistemic authority about it? Why is history important enough to require us to regularly decide whose account to believe? In Chapter 5, I will aim to answer this question. I will argue that the past is special for human beings because humans do not only represent the physical but also the social effects of the events unfolding around them. In contrast to the physical consequences of a given event, its social consequences are often not perceivable and consist only in mental

representations. Therefore, when trying to prove their existence, there is often no other way than to point to the event that established a given social effect in the first place.

Overall, this thesis will give an account of the cognitive architecture allowing us to think about events in a quasi-perceptual manner (Chapter 2), how episodic memory results from and goes beyond this architecture and how this gives rise to our intuitions of what 'remembering' is (Chapter 3), why humans have evolved the ability to 'remember' in this way (Chapter 4), and, finally, why remembering the past and the ability to give testimony about it plays such a prominent role in human culture and social life (Chapter 5).

CHAPTER 2

EPISODIC SIMULATION: THE NEURO-COGNITIVE BASIS OF EPISODIC MEMORY

If any agreement about the nature of episodic memory exists, it is about the fact that episodic memories are representations of specific, past events. A lot of psychological research on episodic memory in recent years has, therefore, not necessarily focused on our ability to remember the past per se. Instead, spurred by a series of exciting results in neuropsychology and cognitive neuroscience, researchers have focused on our ability to think about events that we do not currently perceive. In this investigation, it has become apparent that the best way to think about episodic memory from a neuro-cognitive perspective seems to be as just one expression of a wider capacity for 'simulating' the perception of events. Episodic memory is thus closely related to other capacities such as thinking about the future or imagining counterfactuals. This finding has induced a lot of debate in both philosophy (e.g. Michaelian 2016) and psychology (e.g. Addis, 2018; Schacter et al., 2012) about the relation between remembering and imagining.

In order to understand episodic memory, it is therefore necessary to first understand this more general neuro-cognitive mechanism for 'episodic simulation'. Thus, in this chapter, I will review some of the evidence for the idea that episodic memory is just one expression of episodic simulation and give an account of the different kinds of representations this capacity seems to be able to produce. Episodic simulation produces mental imagery by constructively recombining elements of past experiences to simulate event representations. However, if episodic simulation indeed produces mental imagery, it remains unclear how the non-depictive aspects of its outputs become cognitively determined. I will argue that there

are (at least) four such non-depictive 'dimensions' of episodic simulation: specificity, temporal orientation, subjectivity and factuality. Further, I propose an account of the mechanisms which might be responsible for determining where a given output of episodic simulation falls within this dimensional space. According to this view, episodic simulation relies on propositional 'scope-operators' in a metarepresentational format, either deployed as inputs to the simulation process itself or produced by post-hoc monitoring processes operating over its outputs. This view will allow us to understand what the contents of episodic memory – as one output of episodic simulation – are. In the next chapter, I will then focus on episodic memory more specifically and ask what might distinguish it from other types of episodic simulation.

1. WHAT IS EPISODIC SIMULATION?

Humans are extraordinarily adept in thinking about circumstances which they do not presently perceive: we can think about *past* events, imagine *future* events, think about how circumstances *could be differen*t from how they are, think about how certain *kinds* of events unfold and even how events are perceived by *someone else*. At first blush, it seems that the all of these abilities must rely on cognitive mechanisms that are quite different from each other. After all, when I think about my last visit to the Zoo, I do not take myself to be thinking about how a visit to the Zoo usually unfolds or what it would be like if I went to the Zoo tomorrow.

Nonetheless, a host of evidence from cognitive psychology, cognitive neuroscience and neuropsychology has given credence to the claim that all of these capacities are very similar in terms of their underlying neural implementation, the cognitive mechanisms involved and

even their phenomenology (Addis, 2018; Buckner & Caroll, 2007; Cheng, Werning & Suddendorf, 2016; Hassabis & Maguire, 2007; Hassabis et al., 2007; Suddendorf & Corballis, 2007; Schacter & Addis, 2007).

On the side of cognitive psychology, it has been found that an 'episodic specificity induction' in which participants are asked to remember a short video clip in as much detail as possible leads to downstream effects in other tasks compared to a semantic control task. For example, after having gone through such an episodic specificity induction, participants generate more details in imagining the future (Madore, Gaesser, & Schacter, 2014; Madore & Schacter, 2015), a higher number of alternatives in which a future event could play out (Jing, Madore & Schacter, 2017), a higher number of intermediate steps in a problem solving task (Jing, Madore & Schacter, 2016; Madore & Schacter, 2014), and more alternative uses for everyday objects (Madore, Addis, & Schacter, 2017). Crucially, these effects do not seem to be specific to an induction in which participants are asked to remember an event. Similar effects have been observed in an 'imagination induction', in which participants are asked to imagine an event (Madore, Jing, & Schacter, 2018). In neuropsychology, it has been found that patients with hippocampal lesions often lose not only the ability for episodic memory, but also the ability to imagine their personal future (e.g. Klein, Loctus, Kihlstrom, 2002) and to imagine counterfactual scenarios (e.g. Hassabis et al. 2007). Finally, functional neuroimaging studies have shown the activation of a common brain network when participants were engaged in past- or future-oriented mental time travel (Addis et al. 2007; Okuda et al. 2003; for review see e.g. Addis, 2018).

From these and other studies, it can be concluded that there seems to be a neuro-cognitive system which underlies the ability to simulate events. This capacity is subserved by a mechanism located in the medial temporal lobes, centered on the hippocampus, including

fronto-parietal areas and roughly equivalent to the brain's 'default network' (Benoit & Schacter, 2015; Buckner, Andrews-Hanna & Schacter, 2008; Schacter et al., 2012; Spreng et al., 2008). The precise function of the hippocampus and surrounding areas is contested (Kim et al., 2015) and different theories describe the hippocampus' role in the storage, consolidation and retrieval of information (McClelland, Naughton, & O'Reilly, 1995; Moscovitch & Nadel, 1998; Nadel & Moscovitch, 1997; Squire, 1992a). Nonetheless, there is now substantial evidence that the hippocampus serves a function that is not solely mnemonic (Eichenbaum & Cohen, 2014; Maguire & Mullaly, 2013; Maguire, Intraub, & Mullally, 2015; Schacter & Addis, 2012). Specifically, it has been shown that the hippocampus plays a role in spatial navigation (Spiers & Maguire, 2006), counterfactual and fictitious thinking (De Brigard et al., 2013; Addis et al., 2009; Schacter et al., 2015) as well as episodic future thought (Addis, et al., 2009, Addis, Wong & Schacter, 2007).

1.1. THE MECHANISMS OF EPISODIC SIMULATION

Originally, the mechanism underlying these capacities was characterized by various authors as being constrained to enabling 'mental time travel' (Suddendorf & Corballis, 2007; Schacter & Addis, 2007; Tulving, 1983; 2002a); that is, the simulation mechanism was thought to enable the mental simulation of events situated in subjective time (Tulving, 2002b). Alternatively, Hassabis & Maguire (2007; 2009) proposed that the function of this mechanism should be characterized in terms of a 'construction system' allowing the simulation of 'scenes'. Hassabis and Maguire argued that the distinctive feature of the outputs of this system was not their temporal orientation but rather their *spatial coherence*. It has since been recognized, however, that neither temporal orientation nor spatial coherence uniquely define the outputs of this simulation system. Rather, it has proven

useful to characterize the function of this network as an 'episodic simulation system' enabling the simulation of event representations more generally irrespective of where in time they occur or indeed whether they occur in time at all (Addis, 2018; Eichenbaum & Cohen, 2014; Cheng et al., 2016; Schacter et al., 2012; Roberts, Schacter & Addis, 2017).

On this view, the episodic simulation system constructs temporally and spatially extended event representations, which *can be* (but are not necessarily) situated in subjective time (i.e. in 'the past' or 'the future'). According to the 'constructive episodic simulation hypothesis' (Schacter & Addis, 2007), episodic simulation functions by retrieving and flexibly recombining elements of stored information. The information drawn on in this process comes from the memory trace, on the one hand, and relevant semantic information on the other (Cheng et al. 2016).

Depending on the functional role a given construction will play, the construction process will then rely more or less heavily on the memory trace or semantic information. For example, the construction of a counterfactual or future-oriented scenario should rely less heavily on trace as compared to semantic information. For example, patients with semantic dementia have been found to be impaired in constructing event simulations about the future (Irish et al. 2012).

Once retrieved, spatial and temporal relations between different elements of stored information represented in the memory trace have to be inferred (Roberts et al., 2017; Keven, 2016). The outcome of these processes is then a quasi-perceptual representation including spatio-temporal information such that predicates like 'next to' and 'before' or 'after' can be applied to its elements.

It needs to be emphasized that episodic simulations are not identical to what has traditionally been called 'memory traces'. For the longest time, and likely due to the fact that our every-day notion of remembering requires it (see Chapter 1), memory traces have been postulated to explain how it is possible for us to remember in the present something that is past. Most traditional thought about memory has assumed that somewhere in the mind there must be stored a structural analog of a past event which we can call upon in remembering it. This idea of memory traces is, however, no longer tenable in view of how remembering (and episodic simulation more generally) seems to operate. Instead, the most plausible view of memory traces now takes them to be distributed neural assemblies, encoding ensembles of sensory features (Robins, 2016). Moreover, one should use the term 'encoding' carefully here. According to DeBrigard (2014b, p.411) "encoding does not seem to lead to any kind of brain modification that remains dedicated to coding for the stimulus that caused it". Instead, memory traces might be best understood as dispositional properties of brain networks to reinstate, in response to the right cue, a specific pattern of activity. As such, memory traces are likely not representational themselves.

2. A PUZZLE: MENTAL IMAGERY AND THE NON-DEPICTIVE ELEMENTS OF EPISODIC SIMULATION

The evidence reviewed above suggests that our ability to think about the future and the past, to think about counterfactuals, and to think about other's perspectives in quasi-perceptual terms can be re-described as aspects of the same capacity. This capacity, roughly, seems to be the ability to simulate circumstances that are not presently perceptually available to us in a quasi-perceptual manner based on past experiences. While this view

makes intelligible from a neuro-cognitive perspective that there might be a unitary capacity of episodic simulation in the human mind, this idea might also be highly intuitive to common sense. After all, it does seem to us that we have the capacity to think in a quasi-perceptual manner about all kinds of circumstances which we do not presently perceive. How else should this capacity function if not by drawing on prior perceptual experiences?

The view that our ability to simulate events is a unitary capacity therefore explains the many similarities that different aspects of mental event simulations display. However, it creates a new puzzle: If all of the different aspects of our ability to simulate events are outcomes of one and the same neuro-cognitive system, how do their differences come about?

Consider that episodic simulation produces the same *kind* of content irrespective of the kind of representation we ultimately construct: Event simulations including spatio-temporal structure (e.g. MacDonald et al., 2011, O'Keefe & Nadel, 1978) and sensory-like information (e.g. Wheeler, 2000), or, in other words, 'mental imagery' (Pearson et al., 2015). Without going into the debate about the relation between mental images and perception, there is substantial evidence suggesting that such images are *depictive representations* playing a functional role similar to perception (e.g. Laeng & Sulutvedt, 2014; Pearson, Clifford & Tong, 2008). That is, mental imagery does not represent information propositionally but rather in a pictorial, depictive format (Pearson & Kosslyn, 2015).

Crucially, however, if the content of episodic simulations consists in such pictorial (i.e. non-symbolic) representations, certain aspects of its outputs remain underspecified. For example, whether a given simulation is about the future or the past can not be depicted. Thus, memories of the past and imaginations of the future are 'independent of their contents' in the following sense: While both will commonly clearly include different

elements, whether the simulation in question refers to the future or the past is underspecified by the mental imagery that has been simulated. The puzzle I want to address here then is the following: *If episodic simulation produces mental imagery and mental imagery is depictive in nature, how are the non-depictive aspects of the outputs of episodic simulation determined?* The fact that in principle nothing seems to differentiate simulations of the past and future in 'what is being simulated' alone makes it necessary to give an account of how our cognitive system distinguishes between these representations. Episodic simulation therefore requires a mechanism that could distinguish between different kinds of simulations independently of their contents (i.e. mental imagery) alone.

This puzzle bears some similarity to what Michaelian (2016) has called the 'process problem', namely "how subjects distinguish between the form of imagination dedicated to reconstructing past episodes – that is, remembering – and other forms of episodic imagination. These include both forms of non-past-directed imagination (e.g. episodic future thought) and forms of past-directed imagination other than remembering (e.g. episodic counterfactuals)" (p. 175). While Michaelian phrases the problem in terms distinguishing between remembering and imagining, I take the problem to be independent of these attributions. From the perspective proposed here, the puzzle consists in distinguishing between different kinds of episodic simulation.

3. THE DIMENSIONS OF EPISODIC SIMULATION

In order to gain traction with an account of a solution to this problem, however, we first need to identify the ways in which episodic simulations can differ in the relevant (i.e. non-depictive) ways. That is, what are the dimensions along which such simulations can differ

which are – in principle – independent of their contents in the sense described above? While other authors have noted the different functions for which episodic simulation can be deployed, there has not been a systematic mapping along which dimensions its output can principally differ. For example, Buckner & Caroll (2007), distinguished different forms of 'self-projection' only according to which cognitive capacity they are in involved in (e.g. prospection, remembering, theory of mind). Similarly, Hassabis & Maguire (2007) distinguished different component processes underlying different possible cognitive functions related to episodic simulation (or as they would call it 'scene construction'). Here, I will aim to give a (possibly non-exhaustive) account of the ways in which the outputs of episodic simulation can differ without relying on identifying the cognitive functions this system is involved in first.[2]

I will propose four such 'dimensions':

- Specificity
- Temporal orientation
- Factuality
- Subjectivity

These are what I will call the 'dimensions of episodic simulation'. The dimensions of episodic simulation can be pictured to open a theoretical space in which we can then locate any possible simulation according to its representational features. This proposal should be understood as a heuristic to enable us to think productively about the space of possible

[2] The proposal which comes closest to mine is that of Michaelian (2016), which, however, lacks a dimension of specificity.

simulations. Having a proposal of what such a space could look like might allow us to explore borderline cases, which might in turn lead to new insights into the structure of this space.

This in turn might result in a taxonomy that could prove productive to empirical investigation of our simulative faculty by exploring the ways in which its different outputs might be produced, impaired and related. Ideally, the ways in which episodic simulation is involved in the cognitive functions identified by other authors will become intelligible simply by thinking about what kind of simulation these functions utilize.

Note that my proposal here is different from other attempts at generating related taxonomies. For example, Szpunar et al. (2014) proposed a taxonomy of future directed cognition which was intended to be independent of the cognitive system each taxon depends on. Thus, their proposal is somewhat orthogonal to the one developed here. Nonetheless, my analysis is of course relevant to future-directed cognition, too, simply because temporal orientation is one of the dimensions I propose. In fact, my proposal can be seen as being complementary to Szpunar et al (2014) in so far as it might provide a starting point in answering the question of what kinds of future thinking depend on the event simulation mechanism described above. My proposal is also intended to be largely independent of classical taxonomies of memory (e.g. Werning & Cheng, 2017; Squire, 2004; Tulving, 1983; see Chapter 1). That is, the proposed dimensions apply to 'memory' only in so far as it is an output of the episodic simulation system. The question of how episodic memory relates to these outputs will be the subject of Chapter 3.

3.1. THE DIMENSION OF SPECIFICITY

The dimension of specificity (sometimes also referred to as 'particularity' or 'singularity'; e.g. Debus, 2014) determines the degree to which a given event representation pertains to one unique, particular spatio-temporal context. In other words, simulating an event as 'specific' results in the representation of this event as having occurred uniquely at one particular point in time. Accordingly, the dimension of specificity should be understood as ranging from 'specific' to 'general'. An example of a specific event simulation would be your memory of how you were riding your bike to work *this morning*. An example of a general event simulation would be your imagination of *the general outline* of your way to work.

Sometimes, this domain has been equated with the distinction between episodic and semantic representations themselves. For example, Szupnar et al. (2014) included a dimension of 'episodic/semantic' in their taxonomy of future thought that seems to mirror what I have called specificity. It is important, however, to distinguish between episodic and semantic event representations on the one hand and specific and general event representations on the other as both general and specific events can be represented either episodically or semantically. For example, you might simply know a description of your way to work without being able to simulate it in quasi-perceptual/episodic terms. Of course, semantic information is likely to play a more prominent role in the construction of a general episodic event simulation (see e.g. Alba & Hasher, 1983; Irish et al., 2012). Nonetheless, since we are concerned with the kinds of outputs the episodic event construction system can generate, phrasing this dimension in terms of specificity seems preferable.[3]

[3] Specificity is closely related to the fact that event simulations seem to include spatio-temporal context; that is, that predicates like 'next to' and 'after' can be applied to their elements. Note, however, that the temporal extension of the simulation does not seem to be mandatory. It seems perfectly possible to simulate a spatially but not temporally extended 'scene' (Hassabis et al., 2007). The same cannot be said for the reverse

Why is specificity in this sense independent of the content of a given simulation? The reason is simply that even the simulation of a 'general' circumstance is always concrete. In order to simulate an event, general or specific, one will always have to simulate a concrete, particular instance of it. The generality of a given event simulation is always 'virtual' in so far as what is being simulated is always a representation of a specified spatio-temporal context. I might imagine the outline of my way to work 'in general' but the simulation in question surely is to some extent specific in its spatio-temporal outline. There must nonetheless be a crucial difference between a representation that is taken to pertain to a general as opposed to a specific circumstance; a difference which is independent of the specificity of the spatio-temporal context included in the simulation.

One might object that, the specificity of a given simulation might simply be given by the amount of detail included in the simulation: It might be obvious whether the simulation in question represents my last trip to work or a general schema of my trip to work simply because my memory of my last trip to work will include more details than a general schema would. It has been argued, for example, that there is a functional segregation within the hippocampus along its long axis related to the resolution of the processed information (Sheldon & Levine 2016; Robin & Moscovitch 2017; Poppenk et al., 2013). While the anterior hippocampus seems to process coarser information such as schemas, semantic concepts and larger scale spatial surroundings, the posterior hippocampus seems to be responsible for processing finer-grained perceptual detail and temporal information. Relatedly, Addis et al.

---

case of a temporally but not spatially extended simulation. Arguably, partly for this reason Hassabis and Maguire argued for the priority of spatial information in the construction process. On their view, the MTL-system should be conceived of as a 'scene' construction system. Whether we should apply the label 'event' to this kind of simulation is, of course, a terminological issue (see e.g. Keven, 2016). And to what extent spatial information is indeed prioritized in or defining for event simulation remains debated (see e.g. Roberts et al., 2017). Crucially, however, specificity is independent of whether or not a given event representation will have temporal extension or not, which is exactly an outcome of the fact that it is not dependent on the content of the simulation.

(2011) found that increased activity in the right anterior hippocampus when participants were asked to simulate specific future events compared to general future as well as specific and general past events. They interpreted this difference as being related to the fact that specific future events are commonly more novel than other types of event simulations.

However, while there seem to be neuro-cognitive differences between specific and general forms of simulations, these differences are merely differences of degree and, moreover, can only be discerned by averaging over a large number of instances. And while it might be true that memories of specific instances will commonly be more detailed than representations of general events, this is certainly not necessarily so. Consider that I might indeed simulate the general outline of my way to work by simulating the last time I went to work (a specific instance of it). Crucially, while the content of the simulation will be a memory of my last trip to work, I might nonetheless represent this simulation as a general schema of my way to work. This in itself showcases the curious dissociation between content and specificity of the simulation: the exact same simulation can be taken to be both a representation of a specific instance *or* a general schema. In practice, it might be the case that we decide whether a given simulation is specific or general by somehow estimating how detailed it is (see below). This, however, will only be a solution to the problem if we assume a mechanism to be in place, which takes the contents of the constructed simulation as input and (by estimating detail) produces a judgment about its specificity. And in this case, too, this judgment would be produced post-hoc and not be part of the contents of the simulation themselves. [4]

[4] Note that there are problems with this solution, too. After all, the puzzle, put in another way, does not consist in 'counting' the number of details of a given simulation but rather in deciding what those details are *details of*.

3.2. THE DIMENSION OF TEMPORAL ORIENTATION

As described above, temporal orientation refers to whether the simulation in question pertains to the past, the present or the future. Your memory of this morning's way to work would therefore be an example of a simulation which would fall on the 'past' spectrum of this dimension whereas your imagination of what will happen on your way to work tomorrow would be an example of a future-oriented simulation. Of course, these representations can be closer or further away from the present moment and accordingly located on each dimension. Notably, one can simulate the present moment as well (think about what is going on in your basement right now) and imagine events without assigning any temporal 'location' to them at all (what would a Quidditch match look like?).[5]

But is temporal orientation indeed independent of content in the relevant sense? A phenomenon that is relevant to answering this question is so-called 'recasting' of memories of the past as imaginations of the future. In principle, when imagining the future, we could simply take a relevant memory of the past and 'recast' it to represent the future. If, for example, you want to imagine what your grandparents' kitchen will look like when visiting them this weekend, you could simply take your memory of your last visit and recast it in this way. The possibility of this phenomenon seems to show that the temporal orientation of a given simulation must be independent of their contents. After all, the same simulation could be assigned past or future orientation without any change in contents.

One might think, however, that recasting includes the construction of two representations instead of merely just the assignment of a new temporal orientation: One might first recall the last visit to one's grandparents and then – from this – construct a second representation

[5] In fact, Addis (2018) argues that episodic simulation is routinely involved in on-line event perception as well.

of the future state of their kitchen. While this is certainly possible, the argument for the independence of temporal orientation from content merely relies on the *possibility* of recasting. Moreover, in practice, we seem to neither commonly engage in recasting (Addis et al., 2009) nor in the construction of two simulations. Instead, we can simply simulate the future on the basis of what we already know about the past without having to remember it first (see Chapter 4).

Research from cognitive neuroscience has found that there are processes that seem to be specifically more engaged in future- as compared to past-directed simulation (e.g. Addis et al, 2009; 2011): The anterior hippocampus seems to be more involved in simulations of future events as well as selectively sensitive to the amount of detail included in such future simulations (Addis & Schacter, 2008). These differences are commonly interpreted to be due to the additional 'recombinational' requirements of imagining the future (Schacter & Addis, 2009; Gaesser et al., 2013) as well as the fact that future (as compared to past) simulation should rely more heavily on semantic information (Irish et al., 2012). These differences, however, are outcomes of the fact that future simulations will differ in content from simulations of the past on average. Moreover, again, these differences are merely differences in degree: Both future and past directed simulations activate the same overall neural systems. While these studies point to the fact that, generally, future and past simulations differ in content, they do not answer the question of how temporal orientation is determined in the first place given that these representations do not *have to* differ in this way.

The idea of a 'dimension of temporal orientation' has sometimes been discussed in the literature in terms of the concept of 'chronosthesia'. Chronosthesia denotes the sense of subjective time, which according to Tulving (2002b), is thought to distinctively accompany

mental time travel in general and episodic memory specifically. Nyberg et al. (2010) investigated the neural correlates of chronosthesia. They trained participants to imagine taking the exact same walk either in the present, in the past or in the future and compared neural activation patterns under each of these conditions. While they did not find any differences between the non-present (i.e. future and past imaginings) conditions, Nyberg et al. did find preferential engagement of the left parietal cortex as well as left frontal cortex, cerebellum, and thalamus when participants imagined taking a walk in the future or the past compared to when they imagined taking the same walk in the present. The authors interpreted these results as a neural correlate of the sense of subjective time accompanying 'mental time travel' into non-present circumstances. While these results suggest that attributing a non-present temporal orientation to a given simulation relies on a distinct (extra-hippocampal) mechanism, it does not explain how temporal orientation is specified (or indeed whether a separate neural process is responsible for such specification).

Moreover, whether chronosthesia should be equated with the relations made available by the processes underlying the dimension of temporal orientation, depends on what we mean by the claim that 'mental time travel' is accompanied by 'a sense of subjective time'. On the one hand, this claim might be understood to entail that temporal orientation of a given event simulation is determined by means of a concept of time akin to McTaggart's A-series (McTaggart, 1908). In other words, on this view, the temporal orientation of 'past' is determined in relation to one's current evaluation of what is 'present'. In this sense, the temporal orientation of a given event representation (in so far as it has a value on the temporal dimension that is not 'present') could indeed be taken to include chronosthesia.[6]

[6] Note, that event simulations do indeed also include temporal relations which would fit McTaggart's characterization of time in B-series terms: One can readily apply predicates like 'before' and 'after' to elements of the simulation simply in so far as it has temporal extension (see footnote 1).

On the other hand, chronosthesia has sometimes been understood to include a kind of temporal self-reference. On this understanding, event simulations would necessarily be situated in subjective time in so far as a given event would represent 'my future' or 'my past'. In this sense, chronosthesia should not be taken to automatically accompany event representations located on the dimension of temporal orientation. Rather, chronosthesia would be a result of certain values of the dimension of subjectivity, which I will turn to next.[7]

3.3. THE DIMENSION OF SUBJECTIVITY

Subjectivity determines whether the event in question is construed from a certain agent's perspective and if yes, whose. Crucially, 'perspective' here, is not identical to perceptual perspective such as 'first-person' perspective vs. 'third-person' perspective as it has been sometimes used in research on field vs. observer memories (Eich et al., 2009; Nigro & Neisser, 1983; Robinson & Swanson, 1993). Rather, simulated events can pertain more or less to a given agent, that is, can be more or less 'personal'.

Of course, this might to some extent correlate with the extent to which the simulation contains a 'first-person perspective'. A given event simulation might, however, not be personal or 'subjective' to a specified agent at all. That is, it might include imagery pertaining to 'an agent' without the identity of the agent being specified. In this case, the simulation in question would fall on the 'a-personal' side of the dimension of subjectivity. In contrast, the event simulation might be represented as pertaining to the 'self' in which case it would fall on the 'personal' side of the spectrum. An example of a personal simulation would therefore

[7] Note further, that chronosthesia even understood in the sense of temporal self-reference should not be confused with Tulving's concept of autonoesis (see Michaelian, 2016).

be my memory of my way to work this morning, while an example of an a-personal simulation would be my imagination of how the storm of the Bastille unfolded. Further, the 'self' is not the only agent to which subjectivity can pertain. One can just as well insert another (well-specified) agent into the simulation. Vicarious memories as described by Pillemer et al. (2015) would represent such case.

Interestingly, the issue of subjectivity in episodic simulation relates to a large body of literature in research on theory of mind. On the one hand, there have been debates about the extent to which the ability to represent others' mental states rely on a capacity for mental simulation (e.g. Goldman, 2006; Stich & Nichols, 1993). One way to arrive at the contents of another person's mental states, the idea goes, would be to simulate the perceptual inputs available to that person, feed it through one's own inferential system and attribute the resulting beliefs/desires etc. as belonging to the person in question.[8] One way to achieve this would be to simulate quasi-perceptual input as belonging to someone else. This is exactly, I suggest, what the dimension of subjectivity allows us to do.

On the other hand, and relatedly, a large amount of research has focused on the ability of 'perspective taking' as intimately related to our mentalizing capacities. In this literature, the type of perspective taking at issue here (i.e. the ability to simulate perceptual information from another agent's perspective) is commonly termed 'Level-2 perspective taking' and is contrasted with the mere 'Level-1' capacity of computing what others can and cannot see or have perceptual access to. There is now a fair amount of evidence suggesting that adults engage in such Level-2 perspective taking spontaneously and automatically during interaction (e.g. Freundlieb, Kovacs & Sebanz, 2018; Elekes, Varge & Kiraly, 2016). One

[8] Of course, while this process would allow us to arrive at the content of another's mental state, another process would have to be in place to attribute mental states to others in the first place.

question which has exercised perspective taking researchers in particular is how we manage to keep multiple perspectives of the same scene apart in online interaction (e.g. Helming, Strickland & Jacob, 2014). This question, is in fact just the same puzzle that is at issue here.

If we had a solution to how subjectivity is assigned to simulations, the problem of how different perspectives are kept apart would be fairly straight-forwardly solved. On the one hand, in online interaction one's own perspective would simply be supplied by ongoing perceptual input while another's perspective would be supplied by episodic simulation 'tagged' to the agent in question. On the other hand, whenever more than one vicarious perspective has to be simulated (and to the extent that working memory capacities allow this), each can be distinguished by the fact that it would be assigned to a different agent.

The claim that subjectivity indeed constitutes its own dimension of episodic simulation predicts that it should be possible to simulate the exact same event from the perspective of the self or another agent without changing anything in the content of the simulated event itself. Because subjectivity in the sense intended here is independent of perceptual perspective, this should indeed be possible. Perceptual perspective underspecifies subjectivity: A given perceptual perspective, by itself, does not specify *whose* 'subjectivity' is represented. I might, for example, simulate a first-person view from the Eiffel tower onto Paris but this alone does not yet determine whose view I am simulating.

3.4. THE DIMENSION OF FACTUALITY

Factuality determines whether the simulation in question is taken to be a representation of an actual circumstance (i.e. a circumstance that has occurred, is occurring or will occur), a non-actual one (i.e. a fictitious circumstance) or of a counterfactual one (i.e. a circumstance

that could, would have, or will possibly occur given certain assumptions). An example of a simulation of an actual event is my memory of this morning's way to work, while an example of a counterfactual simulation is my imagination of what would happen if I had taken the bus (instead of my usual bike). An example of a non-actual/fictitious simulation is my imagination of the story of little red riding hood.

Counterfactuals in this sense are commonly discussed with reference to the past (e.g. Byrne, 2016)[9] and are described as representations generated by thinking about 'what if…'. Multiple studies have shown that episodic counterfactuals rely on default network activation similar to other forms of episodic simulation (Addis et al., 2009; DeBrigard et al., 2013; Van Hoeck et al., 2012; Van Hoeck, Watson & Barbey, 2015). Factuality is 'independent of content' simply because if one simulates an event, nothing of what is happening in the event itself will be the kind of information determining whether the event is a representation of an actual occurrence or merely of a possible one. Factuality cannot be pictured.

Of course, I might be able to *infer* from the elements of the event that this is something, which is highly unlikely to occur. If I imagine what would happen if a pink elephant came crashing through my ceiling, I might infer that this is hardly possible simply because I know things about the world which tell me that pink elephants commonly do not come crashing through ceilings. But this extra inference would exactly be required *after* the simulation has been constructed and not be part of the simulation itself (assuming this is indeed the solution that our cognitive system employs). Moreover, likelihood cannot be equated with

[9] Byrne (2016) proposed that there are 'fault lines' in the way we imagine counterfactuals: people seem to modify certain aspects of reality rather than others when simulating counterfactuals. To the extent that imagination of counterfactual possibilities is an output of the episodic simulation system, this insight becomes relevant to our understanding of how this system operates.

possibility (even though people seem to confuse the two fairly regularly: Phillips & Cushman, 2017; Shtulman & Carey, 2007).

Nonetheless, DeBrigard et al. (2013) found that the simulation of plausible/likely (past) counterfactuals engages brain regions that are more similar to those engaged in episodic remembering than to those engaged in the simulation of unlikely counterfactuals. Neural processing of counterfactuals might thus differ according to plausibility. This difference in degree, however, is yet again merely a reflection of different processes involved in the generation of different contents rather than different types of outputs. It should also be noted that we can generate counterfactuals in ways that are not dependent on episodic simulation. We can think about 'what if…' also in purely semantic terms. Such 'semantic counterfactuals' have been shown to rely on different patterns of neural activation from episodic counterfactuals (Kulakova et al., 2013; Parikh et al., 2018).

Interestingly, Parikh et al. (2018) recently compared activation associated with both episodic and semantic counterfactuals to activation associated with episodic memory and found a reliable difference between counterfactual as compared to memory conditions in dorsolateral and ventrolateral prefrontal cortex as well as temporal cortex. In principle, this kind of contrast would be suitable to detect possible signatures of processes underlying the assignment of factuality. However, as Parikh et al. also note, the generation of counterfactuals is inherently more taxing on working memory since it always involves the generation of two representations: An actual one and, from that, an alternative, counterfactual one. While it is thus possible that the activation observed by Parikh et al. corresponds to processes specific to the assignment of factuality to a given representation, it is likely that it at least partly reflects the increased working memory load involved in the generation of such representations.

This is arguably, a crucial difference between the simulation of a fictitious and a counterfactual circumstance: fictitious and counterfactual simulations differ in the extent to which their relationship to actual reality is specified. The simulation of a non-actual circumstance does not require the prior representation of an actual circumstance while counterfactual simulation does.

Even though counterfactuals are most commonly discussed with reference to the past, if I am correct in claiming that 'factuality' is its own dimension of episodic simulation it should be possible for us to generate episodic counterfactuals with regard to any temporal orientation. Indeed, it seems possible to imagine alternative ways in which the present could be or the future could turn out. While the factuality in the present seems to function much like the past, future factuals, however, are arguably counterintuitive: when simulating the future, are we not always representing how things *could be*? Are simulations of the future not counterfactual by nature simply because there is no fact of the matter from which to infer the 'counterfactual alternative'; is the future not always merely 'possible'?

The theory presented here, however, takes factuality to be an independent dimension of episodic simulation and thus predicts that future counterfactuals should differ from representations of the 'actual' future. Indeed, we can see how this might be the case by appealing to the difference between predicting and 'entertaining' a future scenario. When asked, what will happen when you wake up tomorrow morning, you will be able to give a somewhat precise and confident answer to this. To be sure, when 'predicting' the future in this way, you will know that things 'could turn out differently' and thus assign a likelihood value to your imagination (but see e.g. Teglas et al., 2007 for a prediction which should approach absolute subjective certainty). Assigning a likelihood to your prediction, however, is not the same as entertaining a future counterfactual. When asked what would happen if

tomorrow morning you woke up transformed into a beetle, the simulation you would construct in order to answer this question would not just be unlikely but would actively regarded as 'merely possible'.

4. SOLVING THE PUZZLE?

So far, I have argued that if humans possess an integrated capacity of episodic simulation we are left with the puzzle of how certain non-depictive aspects of these simulations are cognitively determined. I have argued that there are four 'dimensions of episodic simulation' on which simulations can differ and that where a given simulation falls on these dimensions cannot depend on what perceptual content has been simulated.

Before I continue, it is worth emphasizing that the dimensions I have proposed (see Table 1 for an overview of all dimensions and their respective values) might not be an exhaustive list. There might still be others: it is, for example, conceivable to take not only temporal 'location' but also spatial location to be its own dimension (i.e. where in space something occurred). In fact, there does not seem to be a principled, a priori way to determine whether we have mapped the entire space of possible episodic simulations. In approaching the question of how episodic simulations might differ, I have chosen the approach of thinking about what different representational features can be said to be independent of the mental imagery produced. This approach, however, necessarily involves testing what kind of conscious simulations we can produce in a top-down manner (for a similar strategy see Chater & Oaksford, 2013): if we can produce a simulation differing on a given dimension without differing in the actual imagery, this should count as evidence for the existence of such a dimension.

To drive this point home, and to illustrate the dimensional space generated by my proposal, it might be helpful to produce a few examples of simulations involving possibly counter-intuitive combinations of the dimensions I have proposed. In answer to each question below, it should be possible to generate a simulation conforming to the respective location in the dimensional space.

(1) *An example of a general, future, personal, actual simulation*: What will your way to work look like in the winter time when you cannot use your bike anymore but have to use public transport?

(2) *An example of a specific, present, vicarious, counterfactual simulation*: What would your co-worker do now if she, having acquired a medical degree, found you as a victim of an accident?

(3) *An example of a general, past, a-personal, actual simulation*: How did people print books with a traditional, medieval printing press?

Importantly, the top-down approach I have taken should not lead one to think that the questions I pose here about the dimensions of episodic simulation are dependent on our conscious ability to entertain mental imagery. Episodic simulation can operate entirely independently from consciousness (and from conscious mental imagery, Philips, 2014; Zeman, Dewar, & Della Sala, 2015). All my approach intends to show here is that if we can generate, in a top-down manner, simulations differing along a given dimension, then episodic simulation must be able to produce such simulations. Beyond that, however, the evidence from cognitive neuroscience reviewed above independently suggests that episodic simulation can indeed output representations differing along each of the dimensions I have proposed.

Dimension	*Specificity*	*Temporal orientation*[10]	*Subjectivity*	*Factuality*
Value	Specific	Past	Personal	Fictitious
	General	Future	Vicarious	Counterfactual
	A-specific	Present	A-personal	Actual

Table 1: Overview of the dimensions of episodic simulation including the possible values of each dimension. Note that the different values of a given dimension are not necessarily discrete but can also denote continuous differences as in the case of temporal orientation.

4.1. THE ROLE OF PROPOSITIONAL REPRESENTATIONS

The central question of this chapter, however, is how these dimensions come to be specified in the first place. Another way to put this question would be by thinking about what representational format the contents of episodic simulation might take. To reiterate, the puzzle is the following: if it is indeed the case that the contents of episodic simulation are 'mental images' (Moulton & Kosslyn, 2009) how can the non-depictive aspects of those contents be specified? After all, if mental images have the same kind of content as online perception (e.g. Borst & Kosslyn, 2008; Nanay, 2015; Pearson et al., 2015), they do not themselves contain the right kind of information to determine whether a given simulation is

[10] The way I have listed the possible values of the dimension of temporal orientation here assumes that 'present', in effect, amounts to leaving the value of this dimension unspecified. Of course, it is also possible to represent events without assigning them to any time. In this case, the simulation would simply be regarded to be 'not present' without also representing it as 'future' or 'past'. In the scope-syntax framework introduced below, such a representation would still have to be assigned a scope operator but simply one constraining the inferences that can be drawn from the simulation in question so as to not to apply to current reality. Similar points can be made about the other dimensions (e.g. about the values of 'fictitious' vs. 'actual' on the dimension of factuality).

about a specific instance, the future etc. In other words, episodic simulation produces mental imagery, but mental imagery underdetermines crucial aspects of our simulative faculty.

One way to solve this problem would therefore be to complement the mental imagery produced by episodic simulation by propositional information providing an appropriate description of the simulation in question (see Norman & Bobrow, 1979; Conway & Pleydell-Pierce, 2000). A first step in solving our puzzle, would therefore be for episodic simulations to minimally rely on a propositional description of its contents, which determine their place on the dimensions described above.[11]

Note that a simulation of a present, specific, personal, actual circumstance should arguably be functionally identical to current perception. As Addis (2018) has noted, episodic simulation is likely to contribute to ongoing perception and, in fact, a large range of literature has demonstrated the functional similarity of and interference between mental imagery and perception (e.g. Leang & Sulutvedt, 2014; Pearson, Clifford, & Tong, 2008; Pearson et al., 2015). If it is indeed the case that the contents produced by episodic simulation can in principle be functionally identical to online perception, a mechanism is required for 'decoupling' these representations from current reality.

The most prominent way this can be achieved is through a metarepresentational format (e.g. Cosmides & Tooby, 2000; Leslie, 1978). Specifically, Cosmides & Tooby (2000, p. 60)

[11] Alternatively, the relevant distinctions could also be supplied non-propositionally by so-called 'metacognitive feelings' (e.g. Proust, 2007; Dokic, 2014). This solution, however, would entail that we would have to posit a whole zoo of such feelings: past feelings, future feelings, specificity feelings, generality feelings, factuality feelings, counter-factuality feelings, subjectivity feelings etc. Moreover, it is not clear how the complex status of a given simulation of (say) specific, past, actual would be expressed in metacognitive feeling terms. Are such feelings compositional? Another problem with this solution would be that it would have to give an account of how a given simulation could be assigned to another agent on the dimension of subjectivity in terms of feeling alone. Finally, a feeling-based account will likely be only available to the output solution (see below).

proposed that the mind should contain representational capacities implementing a 'scope syntax', that is, "a system of procedures, operators, relationships, and data-handling formats that regulate the migration of information among subcomponents of the human cognitive architecture." In other words, the mind must contain 'scope operators': representational structures allowing the specification of what a given representation can be treated 'as true of'.

From this perspective, the dimensions of episodic simulation can be understood as scope operators determining the scope of a given simulation. Crucially, in scope syntax terms, a representation without a scope operator will be treated simply as true. In other words, 'truth' would not have to be specified through the assignment of a dedicated scope operator but would simply be specified 'architecturally' by the fact that its scope remains unbounded (cf. Sperber's notion of intuitive beliefs; Sperber, 1997). Applied to the current context, this would mean that the *actuality* of a given event representation might not have to be specified but would rather result from leaving the respective variable in the scope syntax of episodic simulation unspecified.

The most obvious way the scope of a given representation can be specified is through a metarepresentational format: a second-order representation specifying the scope of a first-order representation. In fact, Cosmides & Tooby took episodic simulation to be the central application for scope operators. If this view is right, episodic simulation should rely on a second-order representational format (in the sense of requiring the 'representation of a

representation' without necessarily representing the representational relation itself) in order to specify the values on each dimension (for a similar proposal see Redshaw, 2014). [12]

Note that, on this view, different scope operators could have different developmental/evolutionary trajectories such that it would be possible for children (or other animals) to be able to specify some dimensions without necessarily being able to make use of the entire dimensional space available to human adults. On this view, animals capable of some form of episodic simulation minimally require a scope operator distinguishing representations produced by online perception from those produced by episodic simulation (e.g. Hills & Butterfill, 2015). All other, more complex scope operators might be present in some taxa and absent in others.

Interestingly, what I have said so far suggest that the richness of our imaginative faculty, that is, the fact that we can simulate such a vast array of circumstances that are not perceptually available to us, to some extent must depend on our ability to form propositional representations. That is, on the view I have proposed, we would not be able to distinguish between different types of episodic simulation and therefore lose much of what this capacity allows us to do, without the ability to form complex, propositional thoughts. Moreover, this view would leave room for the idea that each dimension is specified by a dedicated mechanism responsible for description of this aspect of the simulation alone. For example, temporal orientation might be specified by a dedicated 'temporal reasoning'

[12] The term 'metarepresentation' might, in fact, not be appropriate here because strightly speaking, scope operators are not representational themselves but rather simply restrict the inferences a given representation can enter (see Chapter 3 for a discussion of the difference between scope operators and proper metarepresentations).

mechanism as described by Hoerl & McCormack (forthcoming) whereas subjectivity might be assigned by a dedicated 'mentalizing' mechanism (e.g. Carruthers, 2009).

To say that a solution to the puzzle of how the dimensions become specified will likely involve a scope syntax based on propositional descriptions in a metarepresentational format is, however, not yet to give an account of how these descriptions come about. How our cognitive system solves this issue is surely a matter for empirical investigation. Nonetheless, I want to suggest two different ways in which it could be solved in principle, thus allowing us to construct a space of hypotheses to investigate. In particular, I will propose two possible solutions to the question of how an episodic simulation comes to be complemented by propositional information: an 'input' and an 'output' solution.

According to the input solution, the solution to the puzzle is to be found in the inputs whereas the output solution includes mechanisms solving the puzzle by working on the outputs of episodic simulation. As will become clear, each view makes clear predictions about the ways in which episodic simulation should operate. Crucially, both of these solutions are accounts of how the *retrieval* of information in episodic simulation operates. I will have nothing to say of encoding simply because at the encoding stage, episodic simulation does not occur. When I watch a movie, for example, the information I encode might be later used to simulate an alternative way the movie could have ended. Such a simulation, however, does not exist whenever I encode the information about the movie and therefore encoding cannot directly contribute to the solution of the puzzle under discussion here. It should also be said that the two views I propose are not necessarily mutually exclusive. It might, after all, be the case that we have multiple ways in which the dimensions of episodic simulation could be specified, corresponding to different ways of retrieving information in episodic simulation.

4.2. THE INPUT SOLUTION

As already mentioned, one way to view the puzzle is as asking for an account of how event simulations can be supplemented by the appropriate propositional descriptions so as to fix each dimension. If we understand the puzzle in this way, one obvious solution might be that the process of episodic simulation, which constructs the respective event simulations, takes such descriptions as inputs. On this view, episodic simulation would depend on a prior, separate process, which feeds it a propositional representation containing a description of the event to be simulated. This propositional description then would contain information as to whether the event to be simulated is to be taken to be specific, past, actual, personal etc. The episodic simulation process would therefore merely consist in a form of sensory 'enrichment' of this description with quasi-perceptual and spatio-temporal information. That is, the description fixing each dimension would not be part of the content of the simulation itself but instead serve as a sort of 'instruction' as to what information to draw on in the simulation process.

Importantly, this propositional description would not necessarily be dependent on top-down intentions (as proposed e.g. by Urmson, 1967). Instead, a given stimulus would first activate (through association, for example) a propositional description, which would in turn serve as input cue to episodic simulation processes. This can happen entirely independently of conscious intention even though we can clearly intentionally generate such cues for ourselves (or they can be provided through verbal instruction by others). One potential problem with this 'input solution', would be that it would have to give an account of automatic, 'bottom-up' instances of simulation such as cases of 'involuntary remembering' (Berntsen, 2009). Consider the famous Proustian cases: One smells the scent of a freshly

baked brioche and is, in one's mind, immediately and vividly transported back to a past instance of enjoying a brioche. It does not seem here as if one somehow constructs a propositional event description which then is only enriched by a simulative faculty. Rather, this seems to be more akin to a case of pattern completion (or 'direct retrieval', Conway & Pleydell-Pierce, 2000) in which one cue automatically leads to the re-activation of an entire array of stored information. An input account would have to posit that this impression is misleading: In fact, the cue (e.g. the smell of freshly baked brioche) activates a propositional association which then in turn is fed into the episodic simulation system operating in a process of pattern completion over the description it received. This, however, seems entirely possible: In principle, nothing prevents pattern completion to function in response to a propositional cue.

4.3. THE OUTPUT SOLUTION

Alternatively, the relevant description of the event simulation required to fix the relevant dimensions might be supplied at the output stage of the simulation process. On this account, after an event simulation has been constructed by the episodic simulation system, it is fed into post-hoc monitoring mechanisms, which then determine where on each dimension it should fall. This solution would bear some similarities to the so-called source monitoring framework proposed by Johnson and colleagues (Johnson et al., 1993; Mitchell & Johnson, 2009). While the source monitoring framework has only been applied to the case of episodic memory, it might be that something like source monitoring applies to all outputs of episodic simulation. According to Johnson et al., after a given episodic memory has been retrieved, monitoring mechanisms determine in a series of steps the 'source' of the episode. Source here is understood both in terms of whether the episode in question was generated by an

actual past event or merely imagined ('reality monitoring', Johnson & Raye, 1981) and, if it is taken to have actually occurred, at what point in the past the episode originated ('source monitoring'). Reality and source monitoring processes are thought to operate over the features of the retrieved representation (Johnson et al., 1988). For example, in order to decide whether a given episode is remembered or imagined, reality monitoring processes are thought to estimate the amount of 'cognitive operations' that are part of the retrieved representation. Since imagined episodes (compared to remembered episodes) are thought to contain more traces of such cognitive operations on average, reality monitoring could use this feature of the retrieved representation as a cue for whether it has been self-generated or encoded through perception. One strength of the source monitoring account has been that it was extraordinarily productive in explaining a host of memory errors such as imagination inflation (Garry et al., 1996) and other source confusions (Johnson, 1991).

In a similar manner then, it might be decided whether a given simulation is specific, past, actual, personal, etc. On this view, the mental images produced by episodic simulation are 'described' by a separate process after they have been constructed. In fact, Michaelian (2016; see also McDonough & Gallo, 2010) has proposed a solution along just those lines. He hypothesized that so-called 'process monitoring' mechanisms would assign a given simulation along the dimensions proposed here before source monitoring processes would be applied. If this is to be feasible, however, the mental images in question have to include characteristics which would allow for such post-hoc process monitoring to reliably distinguish simulations falling on different places in the dimensional space. It would have to be the case that each of the dimensions discussed above can be determined from the characteristics of the simulation itself at least on average. On the one hand, process monitoring mechanisms might operate over the characteristics of the content of each

simulation (e.g. the amount of detail, vividness etc.). On the other hand, these mechanisms might focus on characteristics of the simulation process itself (e.g. the amount of semantic information utilized, ease of retrieval etc.).

While I have mentioned a few possibilities above (e.g. amount of semantic information or amount of detail included in the simulation), it is not clear whether there are candidate characteristics that would reliably allow for post-hoc description on all dimensions. For example, it seems plausible that future, general and counterfactual event representations will rely more on semantic information in the simulation process compared to past, specific and actual events. While estimating the amount of semantic information in a given simulation would thus allow a post-hoc monitoring process to distinguish, roughly, those representations from each other, it would not allow for more fine-grained distinctions. Moreover, since arguably such characteristics would rarely perfectly distinguish where exactly a given simulation falls on a given dimension, we should expect post-hoc monitoring processes to be occasionally error-prone. This, at least, is to be expected if the analogy to source monitoring processes holds.

In this regard, the input and output solutions make clearly contradictory predictions: on the output solution, people should occasionally take themselves to think about the future, say, when they are actually thinking about the past while this should virtually never be the case on the input solution. Michaelian (2016) mentions as one candidate for such confusions the case of cryptomensia (e.g. Brown & Murphy, 1989), i.e. unconscious plagiarism (or, in other words, mistaking remembering for imagining). Cryptomnesia is, however, not the right kind of error: We are not looking for confusions between remembering and imagining per se but rather between different types of simulation regardless of whether they are taken to be rememberings or imaginings (see Chapter 3). Instead, in order to count as confusions on the

dimensions of episodic simulation subjects should take themselves to episodically think about the past when they are actually simulating the future or are thinking of a general pattern when they take themselves to think of a specific instance. While it is possible that such confusions occur, it is not clear how we could identify them.

McDonough & Gallo (2010; 2011) tried to address this question by asking participants to remember autobiographical events and imagine possible future events. They then tested how well subjects could discriminate between these two types of simulation in a consequent memory task. Because participants committed errors in discriminating between memories of memories and memories of future imaginations, McDonough & Gallo then inferred that participants must have engaged in an error-prone post-hoc monitoring process over the features of past- vs. future-directed simulations. As Michaelian (2016) points out, while these studies are certainly interesting regarding the question of how different types of memories are discriminated, it is not clear how much they can tell us about the question of how the dimensions of imagination are determined for a given simulation. After all, the kind of error we are looking for to count as evidence for a version of the output solution would be one in which participants who are (say) instructed to construct a simulation about the future accidentally construct one about the past.

4.4. ARE THE INPUT AND OUTPUT SOLUTION COMPATIBLE?

It is possible in principle that both the input and output solution are correct: it might be, for example, that episodic simulation normally takes propositional descriptions as input but can also be cued in other ways. Whenever propositional descriptions are not available, post-hoc monitoring processes would then have to determine the character of the output in question.

In fact, neuropsychological research on patients producing confabulatory memories suggests that memory retrieval includes both a 'descriptor' component which specifies the characteristics of the episode to be retrieved and a 'monitoring' component which evaluates the retrieved episode in light of the retrieval target (e.g. Burgess & Shallice, 1996; Dab et al., 1999). Given what I have argued here, it is likely that this model fits not only memory retrieval but episodic simulation in general, and, moreover, that specification of the dimensions of episodic simulation can take place both at the 'descriptor' stage and at the 'monitoring' stage.

Nonetheless, it is worth pointing out one further prediction differentiating between the input and output solutions. Since, according to the output solution, the dimensions of imagination are specified by additional mechanisms which are not part of the simulation process itself, this view should predict differences in neural activation between different forms of simulation (likely in frontal areas outside the medial temporal areas) indicating such post-hoc monitoring. Such activation would, moreover, be predicted to occur later in time than the simulation process itself. The input solution, however, would likely predict such activation prior to the actual simulation process reflecting the specification of the description of the relevant aspects of the simulation. While a variety of differences have been observed between different types of episodic simulation (for review see e.g. Addis, 2018; Schacter et al., 2012), to the best of my knowledge, neural activation potentially corresponding to either of these predictions have not yet been observed. This, however, is likely due to the fact that it is difficult to generate the right kind of contrasts which would isolate differences along a single dimension while keeping simulated content constant across conditions. Nonetheless, in principle, one way to investigate how dimension assignment is

solved, would be to look for the neural signatures of the input and output mechanisms proposed here.

5. Conclusion

In this chapter I have aimed to give an account of how our ability to think about non-occurrent events in a quasi-perceptual way comes about. To do so, I have focused on episodic simulation, the process which underlies the human ability to simulate mental imagery about events. We can now understand that in the dimensional space of episodic simulation, the contents of episodic memory only inhabit one specific point: That of specific, past, personal, actual events. It is however, not clear how different points in this dimensional space are cognitively distinguished from each other. I have proposed two possible solutions for how this puzzle could be solved in principle: either as descriptions supplied as inputs to the simulation process or as descriptions specified by monitoring the features of the outputs of this process. Moreover, it is possible that both mechanisms are at play and might work in tandem with each other.

This gives us an idea about the neuro-cognitive mechanisms that underlie our ability to think in quasi-perceptual terms about past events. Do we therefore also have an account of episodic memory? After all, the taxonomy I have reviewed in Chapter 1 suggests that episodic memory is nothing but our ability to remember past events. Therefore, one might think that episodic memory is just one output of episodic simulation, namely, a simulation of a specific, past, personal, actual event. Episodic memory, however, is more than a mental event simulation. In what way episodic memory differs from such representations will be the topic of the next chapter.

CHAPTER 3

THE CONTENTS AND FORMAT OF EPISODIC MEMORY

In the last chapter, I have argued that humans possess an integrated capacity to simulate in a quasi-perceptual way events which they cannot presently perceive. I have also argued that one of the ways humans can simulate such events is as having *actually* occurred to *themselves* in the *past*, at one *specific* point in time. To get episodic memory, however, I will argue, we need more than that. Therefore, in this chapter, I will zoom in on episodic memory. I will give an answer to the question what we should take remembering to be, how it relates to episodic simulation and other conceptually similar kinds of representations such as semantic memories or beliefs about past events. Finally, the account proposed in this chapter will also offer an explanation for why our everyday notion of remembering is structured in the way described in Chapter 1.

1. WHAT IS EPISODIC MEMORY?

What then is episodic memory? The term 'episodic memory' entered the repertoire of cognitive psychology some time ago, and is often presented as roughly corresponding in function to the use of the word "remembering" (Tulving, 1985; Gardiner, 2001). The fact that we seem to have no trouble identifying instances of remembering in everyday life, however, obscures many cognitive and conceptual subtleties in relation to episodic memory. The term is often used in slightly different ways by authors with differing theoretical

inclinations.[13]

As discussed in Chapter 1, human memory is typically partitioned into separate systems along two axes (Squire, 1992b): declarative/nondeclarative and long-term/short-term. Within this taxonomy, there are two separate declarative, long-term memory systems: semantic memory and episodic memory. Therefore, the effort to understand episodic memory has traditionally focused on identifying those of its features that distinguish it from semantic memory.

Tulving (1972) originally defined episodic memory as memory for personally experienced past events. Episodic memory, in this conception, was thought to uniquely include information about what happened, when, and where (so-called WWW information). However, this kind of information can be represented in semantic memory as well (Klein 2013b): One can, for example, recall this morning's breakfast in terms of WWW information purely by invoking semantic memory. Tulving (1983; 1985; 2002a) thus subsequently amended his definition by adding that episodic memory is distinguishable from semantic memory because of its unique phenomenology. Whereas information in semantic memory is thought to be simply known, episodic memory comes with "mental time travel"; that is, when we remember an event, we re-experience the event as it occurred. Tulving labeled the different phenomenological states of semantic versus episodic memory as "noetic" and "autonoetic" consciousness, respectively.

Partly due to the phenomenological nature of this distinction, much discussion has focused

[13] Specifically, it is not always clear how the distinction between episodic memory and autobiographical memory is drawn. I take autobiographical memory to refer to knowledge about the self and take the fact that episodic amnesiacs do not always lose this kind of knowledge to speak in favor of distinguishing between these concepts (Klein & Gangi, 2010; Klein et al., 1996; Picard et al., 2013). It is my understanding that autobiographical memory is a specific kind of memory content, which can be, but is not necessarily, represented in episodic memory.

on what autonoesis should be taken to be. From this debate, two main lines of thinking have emerged. On the one hand, authors such as Russell and colleagues (Clayton & Russell, 2009; Russell, 2014; Russell & Hanna, 2012; for a similar view, see Hills & Butterfill, 2015) have proposed a minimal characterization of episodic memory. In this view, episodic memories are re-experienced and thus distinguished from semantic memory by the fact that their contents are WWW elements bound together into a holistic representation. That is, because such memories have spatiotemporal structure (such that predicates like "next to," "before," or "after" can be applied to their elements), and include perspectivity as well modality-specific sensory information, they carry all the features of ongoing experience. Further, because such episodic memories would represent completed events, they could be identified as "past" in a minimal, non-conceptual sense (Russell & Hanna, 2012). Autonoesis might then simply be a by-product of the quasi-experiential character of such recalled events.

On the other hand, many have argued that episodic memory includes more than just event information (Dokic, 2001; Klein 2013b; 2014; 2015b; Klein & Nichols, 2012; Perner, 2001; Perner et al. 2007; Perner & Ruffman 1995). In this view, when we remember an episode, we represent more than just the event itself; we further represent that we had personal experience of the event in question. Specifically, Dokic (2001) has argued that we should understand the difference between episodic memory and other types of memory as evidenced by the fact that "genuine episodic memory gives the subject [...] a reason to believe that the information carried by it does not essentially derive from testimony or inference but comes directly from the subject's own past life" (p. 4). Klein and Nichols (2012) supported a similar view in their report of the case of patient RB, who seemed to have lost the capacity to autonoetically remember the past. This patient reported having lost the

capacity to non-reflectively tell "from the first person, 'I had these experiences'" (p. 690). Autonoesis thus seems to carry propositional content to the effect that the information in question was acquired firsthand. To account for this circumstance, such self-reflexive views of autonoesis usually take episodic memory to be metarepresentational.[14] After all, to represent that one's memory is the outcome of a past experience, one has to represent the representational character of the memory itself (Perner 1991).

1.1. THE CONTENTS OF EPISODIC MEMORY

How can we reconcile the view of autonoesis as resulting from the 'quasi-experiential' features of recalled content with the view that autonoesis is metarepresentational? First, we can note that the 'quasi-experiential' character of episodic memory must be due to the fact that its contents are products of the episodic simulation process. In other words, the contents of episodic memory are an episodic simulation about a specific, past, personal, actual event. As such, as I have argued in Chapter 2, these contents must consist in mental imagery and thereby represent perceptual properties. Episodic simulations of specific, past, personal, actual events can therefore qualify as 'minimal' episodic memory: They are quasi-experiential in the sense of including spatiotemporal context, perspectivity, and modality-specific sensory information.

Crucially, episodic memory is therefore not identical with the memory trace on the basis of which a given simulation is constructed. Memory traces do not 'represent' events but rather

[14] Here, I adopt Perner's (1991; 2012) view of a metarepresentation as a "representation of a representation as a representation." This formulation is stronger than conceptualizations proposed by other authors who conceive of a metarepresentation as merely a "representation of a representation" (e.g., Sperber, 2000). However, our characterization of episodic memory as metarepresentational in this strong sense does not entail that all metarepresentations must be of this kind.

allow the episodic simulation system to reinstate perceptual states. That is, episodic memory can be based on a memory trace but is not identical to it. In fact, as we will see below, given the simulative character of its contents, trace information involved in the generation of the contents of a given episodic memory does not have to be 'causally connected' to the simulated event at all. The reason for this is that one can simulate an event that never occurred and still represent it as specific, past, personal and actual (e.g. Loftus & Prickell, 1995). Moreover, episodic memory is therefore also not identical to semantic information as has sometimes been argued by proponents of the self-reflexive view of autonoesis (e.g. Perner & Ruffman, 1995). Semantic information can serve as an input to the simulation process allowing the simulation of circumstances according schemas and scripts (Alba & Hasher, 1983) but has to be distinguished from the outputs of this process.

1.2. THE FORMAT OF EPISODIC MEMORY

Episodic simulation alone is, however, not sufficient for episodic memory to occur: Hippocampus-based simulations become episodic memories only when they are conceptualized in a certain way, namely, as the outcome of past first-person experience. The event simulation itself (even including descriptions on the dimensions discussed in Chapter 2) does not seem to differentiate between imagined and remembered events (see below). Some authors have proposed that autonoesis serves as a 'memory index': a representational tag differentiating episodic memories from imaginations (Klein, 2014; Michaelian, 2016). In this view, the difference between imagined and remembered simulations equates to the difference between factual and counterfactual event representations. Autonoesis would then allow us to differentiate between factual and counterfactual representations.

1.2.1. Distinguishing between remembering and imagining

If the difference between imagined and remembered event representation came down to the difference between factual and counterfactual simulations, the imagine/remember distinction would come for free through the mechanisms specifying the dimensions of episodic simulation discussed in Chapter 2. It might seem plausible, for example, that actual, past, personal simulations will simply be taken to be remembered by default while other types of outputs must be taken to be imagined. In this case, a 'memory index' in the form of autonoesis would not be necessary because no information additional to a given simulations place in the dimensional space would be necessary to distinguish remembered and imagined simulations.

Crucially, however, fixing a given simulation on the dimensions of episodic simulation underspecifies whether it should be taken to be remembered or imagined. To see this, note that both remembering and imagining a specific, past event are compatible with the belief that the event indeed occurred. One can (even accurately) imagine a past event that one believes to have occurred (even to oneself). This is, in fact, common when we represent events of which we have only secondhand information (see also Pillemer et al., 2015). The simple reason for this is that, while remembering can only apply to specific, past, personal, actual event simulations, imagining might apply to *all* outputs of the episodic simulation system including specific, past, actual, personal ones.

It is worth providing an example here. Take an instance of alcohol-induced amnesia: you might not remember that you texted your ex-partner when you were drunk last night. Now your friend who was with you at the time of the texting tells you about this episode. As a

consequence, you form an accurate event simulation based on your friend's account of the events. The resulting representation will now have a lot of the features of episodic memory without being, strictly speaking, a memory: It will be about a specific event in the past, it will include a first-person perspective, it will be vivid, you will believe that the events it represents occurred, and you will also believe that you were the agent involved. Thus, you will have generated a simulation of a specific, past, actual and personal event.

Importantly, however, this representation will not include the crucial piece of source information telling you that you know about the events of last night through your own experience. After all, the source of your knowledge about the events in this case will be your friend's testimony and not your own experience. Thus, it will not be autonoetic in the sense of telling you that you came to know about this event through your own experience. Hence, it seems clear that the bearer of such a representation will not claim to remember. This example illustrates a number of things. First, it is clear here that autonoesis should not be taken to be part of the content of the memory. The event you simulate can (in principle) be the same regardless of whether you formed it on the basis of your friend's testimony or not. Second, we can see that source information is importantly different from agentive information: That is, knowing that "I did X" is not the same as knowing that "I know about X because I experienced it." Therefore, autonoesis cannot be an outcome of specifying a given event simulation on the dimension of subjectivity. 'Personal' event simulations in which oneself is the agent have to be distinguished from event representations acquired through first-hand experience: Even though one commonly implies the other, as the example above

illustrates, they can dissociate.[15]

Moreover, if the content of autonoesis is indeed a proposition to the effect of "I had these experiences," it can not be an outcome of the specification on the dimension of factuality either. While autonoesis might then only describe events which actually occurred, it is not identical to the belief in the occurrence of an event. Whether an output of episodic simulation is about a past, actual, personal event does therefore not yet allow one to determine whether it is imagined or remembered. Thus, the distinction between simulations on different points in the dimensional space does not cleanly distinguish between imagined and remembered simulations. Instead, autonoesis marks those events of which one had firsthand experience as opposed to some other source. Therefore, while autonoesis might indeed be a 'memory index', the remember/imagine distinction applies to the 'mental origin' (i.e. the source) of the simulation in question. When we remember an event, we take our current representation of it to be an outcome of our own first-hand perceptual experience, while in imagination we do not specify its origin in this way.

1.2.2. ATTITUDE ATTRIBUTION

This point applies to other attributions beyond 'imagining' and 'remembering' as well. 'Planning', 'desiring' etc. are similarly not specified just through a given simulation's location in the dimensional space as I have discussed it so far. Whether a given simulation represents, for example, a desired outcome is underspecified by its specificity, temporal orientation,

[15] Of course, if one believes that something occurred in one's own past one will be inclined to also believe that one knows about this through first-hand experience. This heuristic inference might explain a range of memory illusions and false memory effects such as misinformation effect (Loftus, 2005) and implanted memories (Loftus & Prickell, 1995): People tend to assume that the source of their present representation of events that happened to them lies in first-hand experience.

subjectivity or factuality. The simulation itself will not decide whether (say) a belief or a desire is simulated. Instead, this has to be additionally specified.

All of these attributions can be described as 'propositional attitudes', that is, the parts of a representation which determine its functional role over and above its content alone. For example, the attitude of 'believing' would determine a representation to be taken to be 'true of the world' while that of 'desiring' would determine it to be 'a state worth bringing about in the world'. In this way, you can both 'believe' or 'desire' the same content (e.g., that "there is chocolate in the fridge") without changing anything about the content itself.

To understand how attitude ascription might work, we can return to Cosmides and Tooby's (2000; see also Klein et al., 2004) account of scope syntax described in Chapter 2. One way the scope of a given representation and the resulting functional role that a given output of scenario construction ought to play in inference can be determined is by specifying its source. This, in turn, necessitates that the contents of the construction be representationally decoupled from their direct relationship to reality. Indeed, source-monitoring mechanisms seem to fill the role of such decoupling processes; they effectively endorse contents under a given description (Michaelian, 2012a; 2012b).

The decoupling process, Cosmides and Tooby (2000) argued, is best described as the application of an appropriate propositional attitude. In the case of episodic memory, the attitude of '*remembering'* corresponds roughly to describing a given first-order representation as "has been obtained through firsthand informational access." Cosmides and Tooby went on to propose similar attitudes for imagination, planning, and so forth. Of course, attitudes cannot be indiscriminately applied to any content; for example, one cannot remember a future event. However, this proposal makes sense of the fact that the same

simulation of a specific past event can both be remembered and imagined. Moreover, because attitudes can be recursively embedded, this view can accommodate the fact that we can (for example) remember imagining. In effect, the processes involved in assigning the appropriate attitude to a simulation can thus be described as resting on a complex metarepresentational grammar, in which different attitudes, each with their own epistemic status, can be embedded within each other to establish the epistemic status of the construction as a whole.

Crucially, this view preserves the strengths of the minimal view of episodic memory (Russell & Hanna, 2012) in accounting for the distinctive phenomenology involved, while also accommodating the intuition underlying self-reflexive views, according to which episodic content is not enough for episodic memory to occur (Klein, 2013b). Autonoesis is here taken to be an outcome of the capacity to metarepresentationally embed outputs of episodic simulation system under the propositional attitude of remembering.[16]

Given that these 'attitude attributions' seem to be independent of simulated content to some extent, should we take the decision about which attitude to apply to a given simulation to be an additional dimension of episodic simulation? Since some such attitude attributions require the prior specification of other dimensions, it is not entirely clear whether we should count them as a dimension of episodic simulation in the same sense as the ones discussed in Chapter 2. That is, because one can simply not remember the future, whether a given episode can be taken to be 'remembered' would require the prior specification of whether it is of the right kind (i.e. specific, past, actual, personal). Attitude

[16] Note that by using the term 'propositional attitude' here, I do not want to apply that the content that this attitude is applied to has to be propositional itself. "Remembering" (just as other such attitudes) can be applied to non-propositional, depictive representations.

attributions are therefore not 'independent of content' in the same way as the other dimensions discussed here. Nonetheless, it is likely that the processes involved in assigning the dimensions of episodic simulation are also involved in attitude ascription. On the input side, for example, descriptions could contain an additional attitude 'slot' which in turn would govern the functional role of the resulting simulation as 'remembered', 'imagined', etc. On the output side, Johnson & Raye (1981) have described 'reality monitoring' mechanisms which are thought to determine the origin of a given simulation and could therefore fit the role of 'attitude ascriptors'.

2. EVENT MEMORY AND EPISODIC MEMORY

Crucially, however, not all attitude ascriptions are metarepresentational in this way. One exception here is the attitude of 'believing': As mentioned in Chapter 2, in terms of scope syntax, treating a given representation as 'true' can simply result from not assigning to it any scope operator. Such a representation would then simply function as a belief. The attitude of 'believing' might therefore be one case in which leaving a given attitude unspecified on the dimension of factuality would result in it being treated as 'true of the world' and thereby functionally identical to a belief.

Note that *believing* is a propositional attitude as well, albeit one that does not necessarily require the metarepresentation of its contents. Arguably, we similarly adopt the attitude of believing to semantic memories by default: that is, semantic information is often treated as architecturally true in virtue of having an unbounded representational scope. Thus, if the same attitude is applied to a given episodic simulation (i.e. by not assigning a scope operator

on the dimension of factuality to them), we should expect its content to acquire inferential properties similar to semantic information.

Therefore, episodic simulations do not have to be embedded under a metarepresentational attitude in order to support behavioral decisions. This at least is suggested by findings showing that the hippocampus is implicated in implicit memory tasks (Hannula & Greene, 2012; Hannula et al., 2012; Olsen et al., 2012; Sheldon & Moscovitch, 2010). That is, simulations of specific past events can be represented as having occurred without the attitude of remembering being applied. However, although they do not differ in content (they are about specific, past, personal, actual events), believed event simulations are not episodic memories.

Instead, I will reserve the term *event memory* for this kind of representation (for similar proposals, see Keven 2016; Rubin & Umanath, 2015). Such event memories might differ from full-blown episodic memories in that they do not include source information and are not necessarily subject to conscious awareness (Dew & Cabeza, 2011; Hannula & Ranganath, 2009; Henke, 2010; Moscovitch, 2008).

Such a distinction between event and episodic memory is at least tentatively supported by findings from several lines of research. Infants demonstrate some capacity for recalling events (Bauer & Leventon, 2013; Mullally & Maguire, 2014), but only between the ages of three and five years do children begin to access event information as the source of their beliefs (Haigh & Robinson, 2009). Moreover, the outputs of the hippocampus are not necessarily conscious (Henke, 2010), but they nonetheless inform eye-movement behavior in implicit memory tasks (Hannula & Ranganath, 2009). In fact, eye movements can serve as an implicit, veridical index of event memory, which can dissociate from explicit responses

(e.g., Hannula et al., 2012). On the side of neuropsychology, the case of RB mentioned above demonstrates that it is possible to lose the capacity to remember events autonoetically without losing the ability to access event information as such (Klein & Nichols, 2012).

The concept of event memory thus allows us to take seriously the mnemonic abilities of young children (e.g., Burns et al., 2015; Clayton & Russell, 2009; Fivush & Bauer, 2010) and nonhuman animals (e.g., Clayton & Dickinson, 1998; Gupta et al., 2010; Martin-Ordas et al., 2010; 2013; Templer & Hampton, 2013) without necessarily attributing to them capabilities for episodic memory in the same sense as human adults possess them (Redshaw, 2014; Tulving, 2005).

2.1. NARRATIVITY: INFERRING CAUSAL AND TELEOLOGICAL RELATIONS BETWEEN EVENTS

One more point bears emphasizing: In episodic memory, we do not simply generate a single event representation but rather chains of inferentially connected events. By connecting event representations into such chains, we can simulate complex circumstances unfolding over longer time periods and represent overarching connections between events. That is, we can think about not only how occurrences are located relative to one another in time and space ('before'/'after'/'next to') but also how events are connected causally ('X occurred because of Y') and teleologically ('X occurred so as to bring about Y'). The ability of inferring the causal and teleological relations between events, in effect amounts to our ability to 'think narratively'. As Keven (2016) has argued, what distinguishes narratives from a mere collection of events occurring in temporal sequence is the fact that causal and teleological relations between events are being highlighted. In other words, when 'thinking narratively' we do not generate an event sequence such as "I went to the store and then I went to the

dairy products section and then I picked up the milk." Instead, we infer the causal and teleological relations between such events resulting in sequences such as "I went to the store to buy milk because there was none left in the fridge after I had come home from work…". This allows us to 'make sense' of events (Bietti, Tilston & Bengerter, 2018) and relate them to the present.

Episodic memory can therefore be viewed as 'narrative' in character. According Keven (2016; 2018), this narrative character is supplied by a 'narrative binding mechanism' which infers the respective causal and teleological relations between different event representations. Narrative binding seems to be intimately involved in episodic memory development and spared in episodic amnesiacs with bi-lateral medial-temporal lobe damage (Keven et al., 2018). This suggests that narrative binding is independent to some extent from episodic simulation. Nonetheless, 'narrativity' thus understood is likely not unique to episodic memory but can apply to all outputs of episodic simulation. While in the case of past events, narrative binding allows us to understand how the past causally and teleologically relates to the present, narrative binding might also function to support future planning or counterfactual causal thinking. I will return to the role of narrativity in episodic memory in Chapter 5.

3. REMEMBERING AND BELIEVING THE OCCURRENCE OF PAST EVENTS

One consequence of viewing episodic memory as the outcome of the application of a distinctive propositional attitude is thus that remembering has to be distinguished from believing. This might seem counterintuitive because we usually believe whatever we remember. Nonetheless, psychologists commonly distinguish the belief in the occurrence of

an event from episodic memory of the same event (Blank 2009; Fitzgerald & Broadbridge 2013; Mazzoni & Kirsch 2002; Rubin et al. 2003; Scoboria et al. 2014). What, then, should we take the relationship between remembering and believing to be?

3.1. EPISTEMIC GENERATIVITY

Crucially, when we remember, we do not simply generate two representations: a belief that the event in question happened and an episodic memory of the event. Instead, these representations are inferentially connected: We take ourselves to have knowledge about the event in question *because* we had firsthand access to it. Perner and Ruffman (1995), followed by Suddendorf and Corballis (1997; 2007), took this circumstance to imply that episodic memory requires a form of causal understanding: namely, the capacity to understand that informational access leads to knowledge or belief. They tested this idea by investigating whether there is a correlation between children's episodic memory ability and their ability to infer "knowing" from "seeing." Notwithstanding that Perner and Ruffman did indeed find such a correlation, it seems to me that what is involved in episodic memory is not only a capacity to infer knowing from seeing, but also the ability to further represent the sources of one's own present beliefs *as sources* in the first place (Haigh & Robinson, 2009).

As I have argued above, episodic memory in some sense is just a specific type of source memory. When we remember, the content of the memory no longer functions as an event representation but instead as the source of a present belief. Representing the source of a belief requires, but importantly goes beyond, the inferences involved in ascribing knowledge or belief on the basis of informational access. In the latter case, one simply takes note of the fact that a given agent has appropriate informational access to X and, from this

circumstance, infers that she now knows X. From the fact that Anna has looked inside the box, Ahmed infers that she knows what is inside it. In the former case, however, one has to additionally represent the inferential relationship holding between the episode of informational access and the knowledge state. In this case, from the fact that Anna looked inside the box, Ahmed infers not only that she now knows what is inside but also that this is so because she has seen it.

In other words, to represent the source of a given belief requires the representation of the cause on the basis of which the belief was formed. Therefore, in my account, the represented relation between a given past episode of informational access and a given present belief is one of causation. Episodic memory requires the capacity to understand not only that seeing leads to knowing but, further, the explicit representation of the seeing *as the cause* of knowledge.

Crucially, in most circumstances, the cause of a belief simultaneously provides epistemic justification for it. That is, since belief-forming systems such as perception have evolved to function reliably, if a given belief was caused on their basis, they can be taken to be well justified. Representing the cause of a given belief commonly therefore also provides current justification for it.

Another way to frame the distinction between episodic memory, event memory, and semantic memory would thus be according to their respective role in belief formation: In contrast to event memories and semantic memories, episodic memories are not beliefs but, rather, provide grounds for believing. In more technical terms, event memory and semantic memory are *epistemically preservative*: They preserve the original justification of the endorsement of their contents through time. In contrast, episodic memory is *epistemically*

generative[17]: It generates present justification for why we should endorse its contents (Burge, 1993; Dokic, 2001; Matthen, 2010). When we remember a given event, the fact that we remember supports our belief that this event indeed occurred insofar as it provides a reason for this belief (Teroni, 2014; see also Audi, 1995). If you episodically remember that you were walking on the Red Square last August, you believe that this is indeed what you did simply because you remember it. Other types of memory, in this conception, are different exactly because they do not include a justification of their own contents. When we retrieve information non-episodically, we "just know" without also "knowing why we know."[18] Epistemic generativity is therefore an effect of autonoesis: The fact that the source of our current event representation lies in our past, first-hand experience of the event provides justification for the belief in its occurrence.

3.2. TYPES OF METAREPRESENTATION

As I have argued in Chapter 2, the dimensions of episodic simulation likely require a second-order representational format in the form of scope-operators being applied to the contents of a given simulation in order to fix its scope. While event memories do not require the additional ascription of a metarepresentational attitude since they are simply treated as a

[17] The term 'generative' is sometimes used to describe the view that episodic memory's contribution to the formation of knowledge is the production of new belief (content) due to its constructive character (Michaelian, 2011b). This is then commonly contrasted with preservative semantic memory, which merely preserves beliefs formed in the past without adding to their content. This way of framing the distinction is certainly appropriate to describe the differences between event memory and semantic memory. As far as episodic memory is concerned, however, we want to be clear that we adopt the term epistemic generativity to illustrate that only episodic (but neither event nor semantic) memory produces present psychological justification for beliefs on the basis of constructed content.

[18] In contrast, perceptual beliefs, for example, are entirely transparent. The perceptual character of the belief itself only figures in any inferences drawn from such belief in exceptional circumstances (for example, realizing that one is subject to a perceptual illusion). My claim here is that while we often rely on information about past events in a similarly transparent fashion, in episodic memory proper, the representational character of the memory itself plays a part in the inferences we draw from it (see Burge, 1993).

belief, they do, nonetheless, require scope specification on other dimensions (such as specificity and temporal orientation). Thus, event memories might still be said to be metarepresentational in some sense (Redshaw, 2014). What then is the difference between the form of metarepresentation involved in episodic memory proper and that involved in other forms of event simulation?

Note that, in order to fix the dimensions of episodic simulation, a scope operator has to refer to another representation. In this sense it is metarepresentational. Crucially, however, scope operators do not have to represent the fact that their referents are representations. In other words, while scope syntax requires the representation of other representations, it does not require the representation of the representational relation itself. In episodic memory proper, however, the event representation is represented *as a source* of a current belief. In order to represent epistemic sources (i.e. sources for belief), representational relations themselves have to be represented.

Thus, in episodic memory, event representations are represented *as representations*. The difference between event and episodic memory is thus the type of metarepresentation involved: While event memories are meterarepresentational in the sense that they include representations of other representations, in episodic memory, the representational character of the event representation itself is represented (Perner, 1991; Redshaw, 2014; Sperber, 2000). Put differently, in episodic memory (and not event memory) the representational character of the event representation plays a role in the inferences we draw from it (Burge, 1993).

3.3. MEMORY-BELIEF CONGRUENCY

Remembering and believing thus stand in a relation of justification in which the fact that we remember justifies our beliefs about past events. If this is the case, we might expect the contents of episodic memory to be largely veridical so as to provide normatively appropriate, reliable grounds for our beliefs. In particular, we should not expect our beliefs themselves to have any influence on what we remember.

As illustrated by Neisser's (1981) famous case study of the memory of John Dean, the question of what it means for a memory to be veridical is not a straightforward one. Dean, a former counsel to president Richard Nixon during the Watergate affair, provided testimony that was usually in essence correct but contained many (mostly self-serving) incorrect details. In fact, whether a given memory should be described as veridical might depend on the method used for assessing it (Koriat & Goldsmith, 1996). Consequently, although episodic memory is usually reliably veridical under some descriptions, there has also been a long tradition of research pointing out the fallibility of this system. Starting with Bartlett's (1932) classic treatment, an impressive amount of evidence suggests that the construction process on which episodic memory relies is surprisingly error-prone. Both encoding and retrieval processes typically alter information substantially (e.g., Alba & Hasher, 1983; Roediger, 1996; Schacter, 2001). Crucially, one important line of evidence suggests that beliefs play an unexpectedly large role in the construction of episodic memories (Conway, 2005; Ross, 1989; Blank, 2009). In many situations, construction seems to be guided by one's current beliefs about whatever is to be remembered rather than the memory trace itself. If the construction process underlying episodic memory were indeed optimized to support beliefs about actual occurrences, such a trade-off would be unexpected.

Evidence for top-down influences on episodic memory comes from a range of experiments investigating the effects of post hoc manipulation of participants' attitudes, expectations, or appraisals on their memories. It is usually found in these studies that people remember the past inaccurately but congruent with, and supportive of, their newly acquired beliefs. For example, in a study by Henkel & Mather (2007), participants were asked to make a choice between two options, each of which had an equal amount of positive and negative features associated with it. When asked to remember their choice later, however, participants misremembered the features of the options they chose as more positive than they were (see also Benney & Henkel, 2006; Mather & Johnson, 2000; Mather et al., 2000; 2003). Crucially, this shift was dependent on what participants believed they had chosen, irrespective of their actual choice (see also Pärnamets, Hall, & Johansson, 2015). That is, here participants remembered having made a choice they did not actually make (but believed they did) and, additionally, remembered the option they believed they had chosen as having had more positive features than it actually did. In other words, they displayed both memory congruency with the induced belief and a memory distortion supporting this belief.

Similar congruency effects have been found in such diverse domains as memory for emotions (Levine, 1997), attitudes (Goethals & Reckman, 1973; Rodriguez & Strange, 2015), one's own behaviors (Ross et al., 1981; 1983), one's own traits (Santioso et al., 1990), and even one's own clinical symptoms (Merckelbach et al., 2010; 2011). The methods of these studies are diverse, and it is, therefore, unclear to what extent each of these effects is specific to episodic memory. Evidence suggesting such specificity, however, is supplied by research on memory manipulation.

There is an impressive literature showing that it is possible to induce in people vivid, detailed false memories, which are subjectively indistinguishable from accurate recollections

(Lampinen et al., 1997; Payne et al., 1996). People usually create false or altered memories in response to having changed their beliefs about a given event. This in turn is usually the outcome of having been exposed to persuasive communication (Nash et al., 2015). In fact, persuasion is a main factor in the effectiveness of most memory manipulation paradigms (Leding, 2012). This suggests that induced beliefs can guide episodic simulation.

On the basis of evidence about such belief-memory congruency effects, it seems fair to conclude that retrieval has a tendency to confirm prior beliefs rather than to contradict them. Such evidence then is not easily reconcilable with a view that takes episodic memory to be exclusively aimed at reconstructing events in the way they actually occurred. Rather, these studies show that the episodic simulation process seems to just as often be geared toward constructing event representations so as to be consistent with, and supportive of, our prior beliefs. Commonsensically, we would assume episodic memory to be an exclusively belief-forming system. Phenomenologically, it seems to us that we form beliefs about the past on the basis of remembering it, not vice versa. In contrast, research on memory illusions suggests that beliefs about the past and episodic memory are reciprocally interconnected: Sometimes we remember an event because we believe it occurred.[19] And in turn, once we have constructed a memory on the basis of such a belief, the memory itself might serve to strengthen the belief that induced it.

Crucially, this does not mean that episodic memory is not commonly veridical. In fact, the effects of prior beliefs and attitudes on subsequent memory seem to be highly context dependent (e.g., Eagly et al., 2001). Veridicality in episodic memory construction is not an all-or-nothing affair. Instead, retrieval processes seem to aim to strike a balance between

[19] Compare this to perceptual processes where it is a common trope to point out that high-level beliefs do not and should not have any influence on what we perceive (Firestone & Scholl, 2016).

congruency with memory traces, on the one hand, and belief justification on the other. However, such a balancing act is not always possible. In some such cases, then, remembering an event will lead to belief revision, whereas in others, believing that an event occurred will lead to the construction of an event simulation without a corresponding trace.

4. THE FEATURES OF EPISODIC MEMORY

We are now in a position to specify the features of episodic memory that any functional account should be able to account for. Episodic memory consists of a propositional attitude taken toward the simulation of a specific past event, which serves to justify a belief about the occurrence of this event. I am thus in agreement with Klein (2015), who similarly argued that episodic memory is not individuated through its contents alone but rather through the manner in which this content is made available. More formally, episodic memory is

1. Quasi-experiential

   The representation is an outcome of episodic simulation: It includes spatiotemporal structure, perspectivity, and modality-specific sensory information.

2. Event specific

   The representation is specific to a single spatiotemporal context.

3. Factual

   The event in question is represented as having actually occurred.

4. Past-directed

   The event in question is represented as having occurred in the past.

5. Personal

   The event in question is represented as having occurred to one-self.

6. Autonoetic

   The event representation is (meta-)represented as having been obtained firsthand.

7. Epistemically generative

   The memory is not represented as a belief but as providing grounds for believing.

Importantly, I take these features to be individually necessary and jointly sufficient for episodic memory to occur. Thus, since the fact that episodic simulation is congruency prone is not necessary to episodic memory, I did not list it as a separate feature here. Nonetheless, as I will argue in Chapter 4, I take congruency proneness to be a functional property, that is, a feature rather than a bug of this system. Moreover, we can separate this list of features according to which properties pertain to the content versus the format of episodic memory. Whereas features (1) through (5) pertain to the content (and can thus be shared with event memory), (6) through (7) pertain to the format of episodic memory. The differences between the different kinds of memory capacities discussed above are illustrated in Table 2.

Memory type	Quasi-experiential	Event-specific	Past-directed	Factual	Personal	Epistemically generative	Autonoetic
Episodic memory	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Event memory	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**No**	**No**
Semantic memory	**No**	**Not necessarily**	**Not necessarily**	**Yes**	**Not necessarily**	**No**	**No**

Table 2: Different representational features of episodic memory, event memory, and semantic memory.

5. EXPLAINING THE INTUITIVE NOTION OF REMEMBERING

Given the characterization of episodic memory developed in this chapter, can we explain the intuitions giving rise to the everyday notion of remembering described in Chapter 1? To remind you, in Chapter 1 I took the causal theory of memory to present a codification of our intuitive notion of remembering. According to this notion, someone's mental state counts as a 'remembering' of a given event iff,

(1) the event occurred,

(2) the agent now represents the event,

(3) the agent originally experienced the event as it took place,

(4) the agent's current representation of the event sufficiently resembles the agent's original experience of it, and

(5) there is a causal connection between that original experience and the agent's current representation of the event.

Condition (1) clearly falls out of the fact that episodic memory is a factual event simulation. That is, the contents of episodic memory are located on the 'factual' axis of the dimension of factuality in episodic simulation. They therefore present themselves as indeed having occurred. For this reason, it is not surprising that we take the concept of remembering to include factuality, too.

Condition (2) is arguably due to the fact the episodic memory is epistemically generative and therefore metarepresentational. That is, when we remember, we *know that we are now entertaining a memory of a past event*. As mentioned above, in episodic memory the representational character of the memory figures in the inferences we draw from it (Burge, 1993). Arguably, the intuitive notion of remembering requires the *conscious* representation of the event, and this in turn, is likely due to this metarepresentational aspect of remembering.

Conditions (3) and (5) are outcomes of the autonoetic character of episodic memory: When we remember, we take our current mental state to have been caused by our past, first-hand experience of an event. Therefore, we take remembering to require past, first-hand experience and a causal connection between this experience and our current mental state. Since autonoesis is at the core of what distinguishes episodic memory from other related types of representations, it is not surprising that these intuitions about prior experience and causal connection have similarly formed the core of the intuitive notion of remembering.

The fact that in most philosophical accounts of memory, this causal connection has taken the form of a structural analogue of the event is similarly not surprising. After all, we need to

make sense of the fact that (as expressed in condition (4)), episodic memory has quasi-experiential content coupled with a propositional attitude ascribing first-hand knowledge. In other words, episodic simulation produces content that comes with all the features of ongoing experience. Moreover, in episodic memory this content has been determined to refer to a past, specific event. Finally, and crucially, this content is presented as being the outcome of a past, first-hand experience. Thus, one's current representation of the event must resemble this prior, past experience. And the most obvious way to make sense of this circumstance is to assume that the experience has somehow been 'stored' across time in one's mind.

6. CONCLUSION

In this chapter, I have developed an account of the architecture of episodic memory and its relation to episodic simulation. I have argued that episodic memory consists in the ascription of a propositional attitude to certain outputs of episodic simulation. I have also argued that, if a different attitude is ascribed to these outputs (that of 'believing'), we end up with a different type of memory representation, namely, what I have termed 'event memory'. Finally, I have argued that my characterization of episodic memory can make sense of why we have the intuitions about the nature of remembering that are commonly espoused in philosophical discussions about memory.

If my account is correct, then our intuitions about the nature of memory (as codified by Martin & Deutscher's causal theory) are a natural outcome of the operation of the episodic memory system. Episodic memory presents itself as being a memory as described by the causal theory of memory. However, this presentation is misleading: Most of the conditions

which the causal theory identifies as being ontologically necessary for memory are not psychologically necessary for episodic memory to occur (Michaelian, 2016). We can remember events that never occurred, that we did not experience and that we do not have any direct causal connection to. The episodic simulation process which produces the contents of episodic memory can produce event representations irrespective of whether there is an appropriate memory trace connecting its products appropriately to a specific past experience. Moreover, even if we remember an event that actually occurred, there might be little resemblance between the experience of the original event and our memory.

This is not to say that episodic memory is not generally reliable in allowing us to access the past. Nonetheless, our intuitive notion of remembering should not be taken to be descriptive of the operation of episodic memory. Instead, it might be more appropriate to understand it to be a normative notion: It specifies how remembering *should* operate for us to accept the claims made on its basis as justified. Why, however, should this be the case? Why have humans over evolutionary time developed a system producing such normative intuitions? In the next chapter, I will propose a functional account of episodic memory which might explain why it produces the intuitions it does.

CHAPTER 4

WHY DO WE REMEMBER? THE COMMUNICATIVE FUNCTION EPISODIC MEMORY

Humans are obsessed with their own past. A large part of our conscious mental lives is spent reminiscing about past experiences and sharing those experiences with others (Demiray, Mehl & Martin, 2018; Dessalles, 2007b; Rimé et al., 1991). Psychologists have identified the basis of this obsession as originating in episodic memory. Since Endel Tulving (1972) introduced the concept, the idea that human long-term declarative memory can be partitioned into two separate systems – one semantic and one episodic – has become widely accepted across the field. This agreement, however, has done little to clarify more basic questions about the function of the episodic memory system. Traditionally, most memory research has been preoccupied with studying the capabilities of human memory rather than aiming to illuminate its function. Given the centrality and ubiquity of episodic memory in our lives, it is surprising that the question of the "proper function" (Millikan, 1984) of episodic memory has received attention only in recent years (Boyer, 2008; 2009; Conway, 2005; Cosmides & Tooby 2000; Klein et al. 2002; Hoerl & McCormack, 2016; Michaelian, 2016; Suddendorf & Corballis, 1997; 2007).

In this chapter, I will argue that common accounts of episodic memory function have serious shortcomings, and propose an alternative functional analysis. In my view, episodic memory turns out to be crucial to the human capacity to communicate about past events. Although it is commonly acknowledged that episodic memory is both ontogenetically (Nelson, 1993;

Nelson & Fivush, 2004) and phylogenetically (Suddendorf et al., 2009; Dessalles, 2007a) connected to our capacity to communicate about the past, the exact nature of this connection is usually left underspecified. I propose that episodic memory is essential to managing our discursive commitments by demarcating the range of beliefs about which we can claim epistemic authority. The capacity to manage such commitments in turn contributes to the stabilization of human communication: By taking responsibility for the truth of an assertion (which comes at potential costs), speakers can provide reasons for listeners to believe them. Most importantly, this account can make sense of why episodic memory should be autonoetic – a question that has been left unresolved in the literature so far. Moreover, this account can make sense of a range of empirical phenomena that are not obviously reconcilable with competing explanations.

Overall, my strategy is to reason from form to function: From the design features of the episodic memory system identified in Chapter 3, I infer the cognitive tasks this system has likely been selected to solve. Nonetheless, my account does not make any claims as to the actual evolutionary history of episodic memory, and it addresses only the mature system as it operates in human adults.

Adaptive function cannot be discerned by merely asking what a given cognitive ability is useful for (Millikan, 1984; Sperber & Hirschfeld, 2004): One can use a pair of scissors as a paperweight, but that does not allow one to infer that scissors are designed for keeping paper from flying away. Rather, to arrive at an estimation of proper function, one needs to identify a fitness-relevant problem, which the mechanism under consideration will solve more efficiently than comparable, 'cheaper' alternatives. This then allows one to infer that the capacity in question has been retained in the selection process because of its differential contribution to the solution of said task.

In Chapter 3, I have argued for a distinction between event and episodic memory. According to this distinction, event memories (at least in humans) are simply episodic simulations of specific, past, actual, personal event representations whereas episodic memories are the result of applying the metarepresentational attitude of 'remembering' to these simulations. Episodic memories are therefore the result of conceptualizing first-order simulations as having their origin in first-hand experience thereby supporting our belief in their occurrence. When asking about the function of 'remembering', the question is, therefore, what fitness-relevant problem is solved by an autonoetic and epistemically generative memory system for past events (episodic memory) that could not be solved by a memory system without these features (event memory)?[20]

Before I can give my own answer to this question, however, I will first review and evaluate the most popular alternative view of the function of episodic memory: the idea that episodic memory evolved in order to support future-directed cognition such as planning and decision making.

1. DID EPISODIC MEMORY EVOLVE TO SUPPORT FUTURE-DIRECTED COGNITION?

Information about the past is adaptively important only insofar as it enables us to make better decisions in the present so as to ensure benefits in the future (Klein et al. 2002a). Some authors have taken this constraint very literally, viewing episodic memory as part of a wider system that has evolved to enable us to mentally travel into the future (Michaelian,

[20] Note that I thus take the primary relevant contrast to be the one to event memory and not to semantic memory. To be sure, there is much to be said about the function of event memory (e.g., Nagy & Orban, 2016), but this will not be our focus here.

2016; Schacter & Addis, 2007; Schacter et al., 2007; Suddendorf & Corballis, 1997; 2007). The proponents of this view deliberately frame their account in terms of mental time travel, as they view the abilities of constructing the personal past and the personal future as two sides of the same cognitive system. In this view, the capacity for episodic memory is just one instantiation of a wider ability to construct scenarios in time, the function of which is taken to be planning for and thinking about the future. In essence, since both episodic memory and episodic future thinking (Atance & O'Neill, 2001; Szpunar, 2010) draw on the same underlying mechanism (episodic simulation), these authors take it that both must have evolved for the same purpose: to imagine the future through constructively making available elements of the past, which can be flexibly recombined in the service of simulation (Schacter & Addis 2007; 2009; Suddendorf & Corballis, 2007). That is, because the future is what determines whether one will live to procreate, this aspect of mental time travel should arguably be what caused humans to retain and develop an episodic simulation system over evolutionary time. Episodic memory, too, must therefore be a product of this evolutionary process.

1.1. MENTAL TIME TRAVEL DOES NOT EXPLAIN THE EVOLUTION OF EPISODIC MEMORY

There are, however, several fundamental problems with this 'mental time travel view'. First, even though the proponents of the mental time travel view frequently write as if their account was an explanation of the emergence of episodic memory, it might, in fact, merely be an account of the evolution of *episodic simulation*. As I have argued in Chapters 2 and 3, episodic memory is not identical to episodic simulation. While episodic memory is based on one specific output of episodic simulation, it goes beyond the outputs of this system in so far as it is a metarepresentational redescription of these contents. When asking about the

function of episodic memory, giving an account of the evolution of episodic simulation is therefore insufficient. Future-oriented mental time travel differs from episodic memory in important, functionally relevant respects. Most prominently, episodic memory is autonoetic while future-oriented mental time travel is not. In fact, as Cheng et al. (2016) have pointed out, autonoesis is not necessary for mental time travel to occur (see also Michaelian, 2015).

Beyond the lack of an explanation for the distinctive metarepresentational structure of episodic memory, however, the mental time travel view also neglects important differences between simulating the past and simulating the future. An account of the function of the ability to think about the future episodically does not obviously also entail an explanation of the ability to think episodically about the past. In other words, when giving a functional account of the dimension of temporal orientation in episodic simulation, we might need separate explanations for each of its values. The mental time travel account suggests that the ability to mentally travel into the future simply entails the ability to travel into the past as well. In this view, the ability to represent tensed mental contents at all (i.e. represent contents along a temporal dimension) will automatically entail the ability to simulate future *as well as* past events (Hoerl & McCormack, forthcoming). Thus, it might be said that the subjective past would be a by-product of representing subjective time at all, which in turn, would be an outcome of a selection process driven by the benefits of imagining the future. The ability to assign episodic simulations to the past would then turn out to be an epiphenomenon of our ability to mentally travel into the future.

There are, however, important and adaptively relevant differences between simulating the past and the future which make it likely that these abilities were to some extent subject to different selection pressures (see Chapter 5). When you remember, for example, that you promised your fiancé to come home at 6pm today, it simply does not have the same

cognitive and behavioral consequences as imagining that you might make such a promise at some point in the future. In sum, it might well be that thinking episodically about the future and the past share many similarities because they operate over the same type of content (i.e., event simulations). This fact alone, however, does not explain why we have the ability to do both.

1.2. MENTAL TIME TRAVEL DOES NOT EXPLAIN THE EVOLUTION OF EPISODIC SIMULATION

Given what I have argued in Chapter 2, we might moreover question whether the mental time travel view can in fact serve as an explanation of the evolution of episodic simulation. In Chapter 2, I gave an account of episodic simulation according to which the temporal orientation of a given simulation is the result of the application of a distinctive metarepresentational scope operator. In this view, scope operators specifying the dimensions of episodic simulation will likely to some extent have separate phylogenetic and ontogenetic trajectories. Thus, from this perspective, the mental time travel account is, in fact, merely an account of the evolution of the dimension of temporal orientation. In other words, the mental time travel view boils down to the claim that the ability to apply a temporal scope operator to the contents of episodic simulation evolved due to the benefits of representing the future episodically. An evaluation of this claim goes beyond the scope of the project I am engaged in here. Suffice it to say, however, that the mental time travel view should neither be taken to be an account of the evolution of episodic memory *nor* of the evolution of episodic simulation but rather of the ability to apply tense to episodic contents.

1.3. MENTAL TIME TRAVEL DOES NOT EXPLAIN THE VERIDICALITY OF EPISODIC MEMORY

The mental time travel view is usually presented as having the advantage of being able to explain the constructive character of episodic memory: Imagining the future requires flexible recombination of stored perceptual information. Given that, in this view, selection of this system has been driven by the future-directed aspect of episodic simulation, the past-directed counterpart must be similarly constructive. This line of reasoning is thereby thought to explain the myriad ways in which our reconstructions of the past are error-prone: selection has simply not optimized this system to represent the past accurately.

A problem with the mental time travel view, however, is that this account of constructiveness leaves us without an explanation for why we should ever be able to reliably and veridically recall past events. If evolutionary selection merely constrained our ability to mentally travel in time insofar as it was useful for simulating the future, remembering the actual past should be accidental. The future is not just a replay of the past, and to assume so would leave us unable to predict events based on new contingencies. I take it that the volatility of the future is exactly why this account is attractive as an explanation of the constructiveness of episodic memory. Episodic memory is, however, also reliably veridical in many cases; a fact that becomes mysterious in this view.

One might posit that veridical recollection of the actual past would be helpful for imagining the future: our simulations of the future could be enhanced if we remembered the past first (Szpunar & McDermott, 2008). Selection then might have ensured veridicality in episodic memory because of the benefits an accurate representation of the past provides for our understanding of the future. To be sure, in order to imagine the future, it is important to retain information learned in the past because this will highly constrain any inference as to what might happen in the future. However, it is not clear what re-experiencing the past

episodically does for simulating the future, or how it would contribute more to future planning than what semantic memory, extracted from past experience, could supply. Episodic memory is not identical to stored information, and mentally traveling back to the past will not itself include any information about the future.

In fact, if past- and future-directed mental time travel operate over the same type of content and merely differ in the temporal orientation they assign to their simulations, it is not clear why one would need the past-directed aspect at all to imagine the future. To see this, note that inferring what might happen in the future on the basis of an episodic memory is not the same as mentally traveling into the future in the sense required here. Suppose that, the last time you were at the swimming pool, there was a long line at the entrance. When planning to go to the swimming pool the next time, you might recollect this fact episodically and infer from this that there will likely be a long line again this time. Future-oriented mental time travel, however, is not the outcome of an explicit inference of this kind. Instead, in this case, when you ask yourself whether you should go to the swimming pool today, you might imagine that there will be a long line. Of course, the reason that this piece of information might be included in your imagination of this scenario might lie in the fact that there was a long line last time you were there, and you might even be able to infer this from your imagination. Crucially, however, there is no need for you to represent this when constructing your future swimming pool scenario.

It is thus telling that past- and future-directed mental time travel can be dissociated in episodic amnesia (Maguire et al., 2010; Schacter et al., 2012). The loss of the capacity for episodic memory alone does not significantly impair people's ability to draw inferences about the future. Episodic amnesiacs are not stuck in time: They understand what the future is (Craver et al. 2014b), can make future-regarding decisions (Craver et al., 2014a), and show

normal discounting of future rewards (Kwan et al., 2012). The claim that we can remember the past only in order to imagine the future, then, seems unlikely to be true.

1.4. Source monitoring as a way to guarantee reliability?

One way to reconcile the claim that episodic simulation evolved to simulate future states of affairs with the fact that episodic memory is nonetheless reliably veridical has been to posit post hoc monitoring systems operating over retrieved content (Michaelian, 2012a; 2012b; 2016). Michaelian has proposed that, because selection has not optimized the construction process to accurately represent the past, such accuracy must be ensured post hoc. Because, in large part, the accuracy of our memories depends on their source, and episodes do not include a source tag specifying their origin, the source has to be inferred by monitoring mechanisms at retrieval. Without such mechanisms, the argument further goes, episodic memory would be too unreliable to be useful. Although this assessment is certainly plausible as an account of how episodic memory serves as source memory, it does little to put worries about its reliability to rest. The questions I have raised about veridicality are not issues about source information but rather about the reliability of the construction process itself.

Further, from an evolutionary perspective, if a mechanism carries out its function unreliably, we should expect selection to act on the workings of this mechanism itself rather than producing an additional, expensive, second-order monitoring process. In fact, it is not clear in general why second-order processes would help if we cannot expect certain first-order processes to be reliable. After all, why should the second-order process be expected to be any more reliable? As Kornblith (2012) has pointed out, the assumption that reflection can serve as a way to ensure the reliability of our first-order beliefs generally leads to an infinite

regress simply because reflection cannot guarantee its own reliability (see also Mercier & Sperber, 2011).

2. THE COMMUNICATIVE FUNCTION OF EPISODIC MEMORY

I now propose a novel account of episodic memory function in two steps. First, in this chapter, I address the format of episodic memory by providing an explanation of its epistemic generativity, autonoetic character, and proneness to belief congruency. In the next chapter, I will then address the question of why such a capacity is required for the representation of specific past events.

As discussed above, I take episodic memory to play a generative role in the formation of our beliefs. To get at the proper function of this capacity, let us first consider why it should be necessary to represent our own sources in support of our beliefs to ourselves. One answer to this question has been provided by Cosmides and Tooby (2000): Sources delineate the conditions under which we should revise the beliefs we formed on their basis. Explicitly representing the source for every piece of endorsed information we hold, however, would be both unfeasible and unnecessary. It would be unfeasible because it would require that we store the causal history of any and all inferences we draw, which would call for indefinite storage and computational capacity. And it would be unnecessary because mechanisms of belief updating can be implemented in a manner for which explicit representation of sources is not required (such as Bayesian belief updating).

Therefore, commonly we simply store the outcome of our inferences and discard the history of the inference itself. However, as Cosmides and Tooby (2000) also pointed out, a domain in which sources are regularly useful is the realm of human communicative interactions.

Humans rely on communicated information to an extraordinary extent. Such reliance, however, comes with challenges that necessitate the development of dedicated cognitive machinery. Part of this machinery is the handling of sources and reasons (Mercier & Sperber 2011; 2017).

Most forms of communication are cooperative and, as such, are subject to the same evolutionary stability constraints as cooperation more generally (Tooby & Cosmides, 1992). Communicative exchange of information is beneficial for speakers insofar as it enables them to influence their listeners' mental state. Conversely, listeners benefit from communication to the extent that they can distinguish reliable from unreliable signals in order to acquire useful information (Dawkins & Krebs, 1978; Krebs & Dawkins, 1984). As such, communication systems can only survive in the selection process if there is a way to ensure that engaging in information exchange remains beneficial for both parties. On the one hand, if there were no way to ensure that signals were reliable (in the face of possible deception and incompetence), then listeners would soon stop paying attention to them. On the other hand, if speakers had no way of influencing their listeners' mental state effectively and to their benefit, they would stop sending messages (Sperber, 2001).

On this basis, Sperber et al. (2010; Mercier, 2017) argued that we should expect humans to have evolved a suite of capacities that let us – as receivers – scrutinize communicated information for its veracity through assessing both its content and its source. The mechanisms allowing us to do this are collectively referred to as *epistemic vigilance*. These capacities are thought to provide us with the means to avoid being misinformed either through an interlocutor's incompetence or deceptive intent. Conversely, speakers should be endowed with capacities allowing them to effectively influence their interlocutors. According to Mercier and Sperber (2011; 2017; Mercier, 2016) one way this capacity

manifests itself is in our ability for reasoning. Reasoning allows us to argue for why others should accept whatever we are claiming by providing reasons for it. Note that this entails that epistemic vigilance and our ability to overcome such vigilance must be reciprocally interconnected. The better listeners are at scrutinizing communicated information, the better we should expect speakers to be at convincing their interlocutors, and vice versa. Reasoning serves both to maximize the persuasive effects of one's message as well as to scrutinize the validity of the content of received messages. Moreover, one way a speaker might maximize the persuasive effect of her message would be to turn her epistemic vigilance against herself so as to simulate the likelihood that an interlocutor would perceive her intended message as valid. When we reason privately, we in effect anticipate having to convince others. This picture suggests that we should be able to produce reasons for our own beliefs and be sensitive to the quality of the reasons others provide for their assertions.

One way in which source information is important in communication is because it can serve as a reason (Mercier and Sperber, 2011, 2017; Mercier, 2016). On the side of the sender, reasons are important insofar as one can supply them to convince an interlocutor who would otherwise not accept what one has to say based on trust alone. Thus, if someone can point to whatever caused them to believe something (their own reasons), this might be good enough for others to believe it, too. On the side of the receiver, one has to be able to tell good reasons from bad ones when deciding what to believe. This also means that the better senders are at giving reasons the better receivers should be at processing and scrutinizing those reasons, and vice versa.

Communication, then, is clearly a domain where having explicit access to sources is indispensable. In virtue of episodic memory's generative role in belief justification, we might therefore expect it to it play a crucial role in enabling certain kinds of justificatory reasoning,

on the one hand, and supporting epistemic vigilance on the other.

2.1. THE NEGOTIATION OF EPISTEMIC AUTHORITY

Reasons, if we are to identify them as such, are metarepresentational. Taking *p* as a reason for *q* requires more than representing *p* and inferring *q* from it: The fact that *p* and *q* stand in a relation of justification must also be represented. Reasoning, then, is the activity of handling inferences in a way that explicitly represents the justificatory relationships holding between different representational contents. Note that it is not essential that a justificatory relationship actually obtains. Rather, what matters is that such a relationship is represented. You might be wrong in taking the fact that (1) you cannot see beyond the horizon to be a reason to believe that (2) the earth is flat. However, this might not stop you from taking (1) to be a reason for (2). According to the argumentative theory of reasoning, the capacity for representing reasons evolved not because it helps us to draw better inferences but to enable us to make others draw the inferences we want them to draw – that is, to convince them, as well as to evaluate others' reasons (Mercier & Sperber, 2011).

As I have argued in Chapter 3, when we remember, we represent to ourselves why we believe certain things about the past. In other words, we represent to ourselves the relationship between the source of our belief and the belief itself. In essence, sources function much like reasons, with one distinct difference: In order for something to count as a source one has to represent the *cause* of a given belief. Representing the cause of a belief can, and usually does, justify this belief. After all, if one formed a given belief on its basis, the cause in question will likely have been a good reason to do so. Thus, sources and reasons often function in closely similar ways. To reiterate, when we remember, we represent the

fact that we had first-hand experience of a past event which caused us to hold a present belief about this event.

But how could the fact that one remembers serve as a reason for others to believe a given assertion? Note that, in cases where minimal mutual trust between interlocutors can be assumed, it is indeed the case that "remembering" is generally taken by others to be a reason for accepting certain claims. Consider the following situation: John and Jenny are on a walk when Jenny expresses that she is worried that they might have left on the oven at home. To this John replies, "Don't worry, I remember that we turned it off." Why should the statement that John remembers here be any more reassuring than simply stating: "Don't worry, we turned it off"? Here, "I remember" serves as a reason for Jenny to accept John's statement just as it serves as a reason for John to indeed believe that the oven was turned off.

Now, clearly remembering does not work as a reason here in the same way as an argument does. Instead, we can get a clearer sense of the work such autonoetic claims do in interlocution by taking a closer look at the pragmatic structure of testimony. Testimony entitles an interlocutor to take whatever is conveyed as true on the authority of the speaker. This entails that by giving testimony, the speaker herself has to take responsibility for the truth of whatever is stated (Brandom, 1983; McMyler, 2007; Turri, 2011). In the case of secondhand testimony, one can defer this responsibility, but only insofar as one can actually access the source of the information in question.

Indeed, Nagel (2015) has recently argued that our propensity to represent the ways in which our epistemic states are grounded through source monitoring relates exactly to this circumstance. She observes that the different sources of belief we intuitively take to hold

epistemic warrant do not regularly coincide with actual differences in reliability: An expert judgment received through testimony, for example, might well be more reliable than what one has concluded on the basis of one's own perception. It thus seems unlikely that source monitoring would serve a purely epistemic function. Instead, Nagel observed, "source monitoring matters when we need to communicate our judgments to others: indeed, even to decide what does and does not need to be conveyed, it matters where our judgments are coming from, and where our evidence is situated, relative to ourselves and our audience" (p. 301). In fact, the ubiquity with which source information is useful in communication has arguably led to its grammaticalization in about one quarter of all known languages as evidential markers (Aikhenvald, 2004; Speas, 2008). The distinction between indirect and direct forms of evidence seems to be common to all evidential systems.

This begins to make sense of the question of why episodic recollection comes with a representation of its own origin. In this view, autonoesis is the capacity that enables us to distinguish between cases in which we can assert epistemic authority for our own testimony and cases in which we cannot. Note that even in cases where one defers to someone else, one will have to take responsibility for the very act of doing so. If Hanna tells you, "Mary told me that Mark was not at the party yesterday," while Hanna does not take epistemic responsibility for whether Mark was indeed at said party, she does take responsibility for the fact that Mary told her that he was not.

One reason for why it is important to monitor which assertions we can commit to in discourse is reputational. If we discursively commit to, and thereby allow others to rely on, the truth of an assertion, we take responsibility for its truth, and thereby put our reputation as a reliable informant on the line. Thus, discursive commitment comes at a potential (direct or reputational) cost in case our message is found to be unreliable. For our interlocutors, the

fact that we are willing to incur such a cost is a reason to believe us. Through this dynamic, as Vullioud et al. (2017) have recently argued, discursive commitment is a way to stabilize communication. Claims to remembering, then, do not offer a way of overcoming skepticism in the same way as argumentation does.[21] Instead, it is an issue of competence: Episodic memory allows us to signal to others that we indeed have epistemic authority on a certain matter, which in turn commits us to our message, and this should cause others to believe us.

In fact, it is hard to see how else one would argue about certain past events. When it comes to the past, sometimes epistemic authority is all we have to go on when deciding what to believe. Indeed, young children preferentially endorse the testimony of informants who had firsthand informational access (Terrier et al., 2016), and Castelain et al. (2015) showed that young Maya children are more likely to endorse the testimony of a source claiming to have epistemic authority ("The hen went this way because I have seen it") over a source giving no reason, even when it conflicted with another cue usually governing such endorsement (power).

Of course, episodic memory is not the only device allowing us to regulate our communicative commitments. Markers of confidence or pragmatic explicitness seem to be others (Vullioud et al. 2017; Mazzarella et al., 2018). Episodic memory is simply the mechanism specifically geared toward regulating communication about past events. Therefore, communicatively negotiating the past often becomes a matter of convincing one's interlocutor that one remembers: that is, that one has epistemic authority on the matter in question. Because remembering is such an effective way of asserting epistemic

[21] To be sure, we frequently rely on epistemic authority in argumentation. Nonetheless, the mechanisms (by which claims to authority and arguments we try to change others' minds) differ.

authority, it might be beneficial to attribute the origins of (at least certain types of) event information to our own experience in situations in which this would be communicatively useful. This might explain some occurrences of the famous misinformation effect (Loftus, 2005). Here, witnesses have been found to persistently over-attribute misleading information acquired about an event after its occurrence (post-event misinformation) to their experience of this event. From the perspective we have proposed here, this might simply be the best way to make use of this information in appearing as a good witness. After all, if the participants in these studies believed the misinformation to be correct (as they seemed to do), they must have experienced the event in this manner, too.

Going further, this analysis also reveals a functional aspect of the fact that episodic recollections are often rich in contextual details. Although event memory should similarly be characterized by the availability of contextual details, these details play a functional role for communicative purposes in episodic memory. When we debate a past event, the fact that we can produce rich, detailed descriptions can serve as evidence for others – as it does for ourselves, too (Johnson & Raye, 1981) – to believe that we are indeed remembering (Bell & Loftus 1988; 1989). The reason for this effect of detail might be that contextual details (1) give one's interlocutor more leverage to detect potential inconsistencies and reduce vagueness (Kraut, 1978), as well as (2) supply information that might potentially be independently verified. For example, information about the location and co-witnesses of an event makes it possible to potentially obtain evidence about the event that is not dependent on the testimony of one's immediate interlocutor. Such independent verification will, in practice, often not be carried out. Instead, it might be enough that an interlocutor is willing to make her account subject to such verification, which is taken as a reason to accept her testimony. Consequentially, contextual elements that, at least potentially, make verification

possible might be more readily available in recollection simply because this information should allow one to be perceived as more convincing. When we argue about the past, we often do not contest whether the event in question happened, but rather in what way it did, and having access to contextual details is often crucial to establish which of multiple accounts of an event should be endorsed and what it should be taken to entail.

2.2. THE CONSEQUENCES OF DISCURSIVE COMMITMENT

Another prediction following from this account concerns the fact that once one has publicly committed to, and therefore taken epistemic responsibility for, the truth of a certain version of events through testimony, this should have subsequent consequences on how and what one remembers. On the one hand, after testimony, it becomes less important to recall the actual event. Instead, to uphold one's commitment, maximize believability, and avoid reputational damage through inconsistency, one should stick to one's own account to a certain extent. In cases where one's account of an event and the actual happenings diverge, one might thus subsequently remember the event in question in a way that supports one's report. A range of memory distortion effects occurring as a consequence of memory report suggests that this is indeed what happens. For example, Cochran et al. (2016) investigated the effect of altering participants' memory reports on their memory for crime events. They found that participants often did not detect the changes to their reports and instead altered their memories to fit the manipulated reports. Tversky and Marsh (2000) found that the public stance one takes on a past event biases recall to emphasize details supporting one's claim (see Higgins & Rholes [1978] and Greene [1981] for related effects). This stance, in turn, has been found to depend on one's particular audience (arguably serving both reputation management and making one's own memory report easier to accept for others),

further altering memory (Pasupathi et al., 1998). In effect, after having reported an event, people often subsequently do not recall the original event but rather a version in line with their latest retelling of it.

The extent to which such distortions would be communicatively useful should be constrained by how skeptical and informed one's audience is. People should be sensitive to the costs of being found wrong, and appropriately adjust the extent to which they prioritize consistency with their own account over accuracy. Thus, the distorting effects of giving testimony might be mediated by how skeptical and informed one perceives one's audience to be. To my knowledge, this prediction has not been tested.

On the other hand, commitment to one's testimony should cause one to be less easily convinced of a different version of occurrences, given that this would undermine one's own epistemic authority. Indeed, participants' susceptibility to social influence has been found to depend on whether they had committed in one way or another to certain details of an event (Bregman & McAllister, 1982; Loftus, 1977; Schooler, Foster, & Loftus, 1988). The reason for this cannot be simply epistemic, because in general, participants have been shown to be quite ready to update their memories on the basis of others' testimony. Instead, my account suggests that participants in these studies became resistant to social influence in order to ensure their own believability.

2.3. Recollective my-side bias

Being able to convince others that we are indeed remembering is only important insofar as it helps us to convince them about *what* we are remembering. The contents of our memories are crucial for supporting certain conclusions over others when it comes to the

interpretation of what a given event entails (see Chapter 5). Thus, if episodic memory indeed has the communicative function of appropriately asserting epistemic authority about the past, we should expect it to make content available in a way that supports our claims.

Mercier and Sperber (2011) have argued that because the production of reasons does not serve normative epistemic goals but is meant to convince others, it should primarily find reasons in favor of whatever we want to claim. Their view predicts the well-known my-side bias in reasoning: the human tendency to reason from conclusions to premises, and not vice versa as normatively required. By analogy, when I claim that episodic memory is crucial for persuading others of a particular version of the past, I should similarly predict such a bias in remembering: to be able to argue for our beliefs about a past event, our recollections should tend to support those beliefs instead of contradicting them.

Indeed, such a recollective my-side bias is instantiated through the way in which our beliefs guide the construction of memory content. Similar to confirmatory reasoning, belief-guided memory construction (reviewed in Chapter 3) can be taken to be a version of the my-side bias to the extent that one constructs a memory justifying what one already believes to have happened. Understanding memory reconstruction as an instance of my-side bias for the purposes of persuasion can make sense of the surprising interplay between beliefs and memory content: the constructive process tends to retrospectively create memories confirming and supporting held beliefs and attitudes. From this perspective, such false memories are simply the results of an inherent tendency to justify our beliefs about the past to ourselves in order to be able to justify them toward others; they illustrate a functional feature, rather than a bug, in the mechanisms of episodic memory. Thus, inducing beliefs about the past in participants is followed by false memories, because once we have

accepted a piece of information, justifiability is ensured through the construction of supporting memory content.

Of course, if I am correct, there should be limits to this form of my-side bias. If the costs of being found wrong are high, or our audience can monitor our assertions effectively, we ourselves should be more skeptical toward the outputs of our own episodic simulation system (i.e., put more effort into checking their consistency) and consequently be less likely to form a false memory.

2.4. Selective remembering and motivated forgetting

A similar analysis can be applied to phenomena described under the heading of "motivated forgetting" (Anderson & Hanslmayr, 2014). Motivated forgetting describes a process by which selective or inhibited retrieval leads to forgetting of aspects of (or entire) events. People tend to selectively remember arguments in favor of an endorsed conclusion or attitude while forgetting counter-arguments against the same conclusion or attitude (Waldum & Sahakyan, 2012). This process has been shown to be especially prevalent in the domain of moral violations. In fact, memories of one's own moral violations are more likely to be forgotten than memories of one's own moral behavior, so that people sometimes seem to display a form of "unethical amnesia" of their past (Kouchaki & Gino, 2016). In contrast, Bell et al. (2014) have shown that memory for the cheating behavior of *others* is well remembered when it is associated with personal costs, but easily forgotten when associated with personal benefits. These processes lead to the phenomenon of rose-colored memories, which emphasize one's own moral character. Given the importance of episodic memory for the communicative negotiation of the past (see Chapter 5), such effects are not

surprising. Both on the individual (Kappes & Crockett, 2016) and the collective level (Coman et al., 2014), selective remembering and motivated forgetting serve communicative ends: convincing oneself simply helps to convince others (von Hippel & Trivers, 2011).

2.5. THE EPISTEMIC VIGILANCE FUNCTIONS OF EPISODIC MEMORY

As mentioned above, epistemic vigilance and the mechanisms designed to disarm such vigilance are essentially two sides of the same coin. The easiest and most effective way to anticipate one's interlocutor's vigilance might be to exercise such vigilance against one's own assertions before uttering them. Source monitoring, as described by Johnson et al. (1993), displays just such a structure. Michaelian (2012a; 2012b) noted that source-monitoring mechanisms are endorsement devices: they decide to what extent we should believe the contents of our own recollections by scrutinizing them for their believability, just as others do when they hear our testimony. These endorsement mechanisms might then be one way in which we can gauge whether we should indeed commit to a certain claim about the past or not. Although Johnson et al. (1993) seem to have assumed that source monitoring is purely epistemic in function and compulsory in event recall, it might well be that these processes are only applied in situations in which scrutiny is required: situations in which one expects to face a (skeptical) audience.

Source monitoring does not just serve anticipating others' vigilance but also functions to exercise vigilance against others. This is borne out by the fact that children become increasingly less suggestible as a result of source memory development (e.g., Bright-Paul et al., 2005; Giles et al., 2002; Lampinen & Smith, 1995). Having access to the sources of our beliefs allows us to keep track of the sources of transmitted information and scrutinize such

sources for their competence and intentions.

Similar to my account, other researchers (Boyer, 2009; Cosmides & Tooby, 2000; Klein et al., 2002; 2009;) have emphasized the role of episodic memory in epistemic vigilance. In their view, the fact that humans so excessively rely on communicated information has necessitated a mechanism allowing us to adjust the truth value of our beliefs according to their source. To decide whether an interlocutor is trustworthy, or whether to re-evaluate such trust, it is necessary to have access to her past behavior in specific situations. When we learn new information about an interlocutor's reliability after the fact, it is important to have access to our interaction history with this specific person to be able to re-evaluate any pieces of information we might have received from her.

The role of source representations in epistemic vigilance is also structurally apparent in how they differ from argumentative reasons. As mentioned above, sources differ from such reasons because they rely on a represented relationship of causation rather than justification between two mental contents. Just because a given belief was caused in a certain way, however, does not necessitate that it is justified. While our belief-forming systems are constituted such that under normal circumstances we tend to form beliefs in a reliable manner, and therefore sources can and usually do justify a given belief, in communication causation and justification can come apart drastically.

The importance of source monitoring in such situations is showcased in misinformation studies, in which participants are able to recover their original event representation when they are informed of the deceptive character of the misinformation (Blank & Launay, 2014). However, as evidenced by the mediocre effectiveness of many such 'post-warnings', episodic memory seems to be rarely used in this way. Most of the time when we are

informed that a given source is untrustworthy, we merely discount this source in the future rather than re-evaluating all beliefs we have form on its basis.

Nonetheless, as predicted by my account, encoding is mediated by epistemic vigilance toward the source of information: misinformation and conformity effects are not automatic but rather depend on participants' evaluation of their own confidence and the reliability of the source of the presented information (Allan et al., 2012; French et al., 2011; Gabbert et al., 2007; Jaeger et al., 2012; Lindsay & Johnson, 1989). When participants have reason to doubt their own (Asefi & Garry, 2003; Clifasefi et al., 2007) or others' ability (Kwong See et al., 2001) or trustworthiness (Dodd & Bradshaw, 1980), they refrain from memory update. In such cases, rather than simply updating their own event representations on the basis of others' testimony, participants encode it in a separate trace (Ludmer et al., 2015).

2.5.1. INTERPERSONAL REALITY MONITORING

The two-sided nature of vigilance and counter-vigilance is illustrated in another aspect of recollection. In deciding whether someone is telling the truth in recounting the past, we usually try to determine whether our interlocutor is remembering or making up the contents of her testimony. Research in the tradition of the source-monitoring framework has investigated how we make this decision about ourselves through so-called reality-monitoring mechanisms (Johnson 1991; Johnson & Raye 1998).

Apart from allowing us to determine whether we should take ourselves to be actually remembering, reality monitoring could play a role in making this decision about others, too. That is, to decide whether we are remembering or imagining a given event, we might use the same mechanisms that are charged with this decision when evaluating others'

testimony. This is suggested to some degree by studies on interpersonal reality monitoring – the ability to judge whether other people's memories reflect real or imagined events (Johnson et al., 1998; Johnson & Suengas, 1989). These studies suggest that participants use the same criteria to evaluate their own memory content and others' memory accounts, and can display above-chance discrimination performance in such situations (Clark-Foos et al., 2015). Note, however, that this is not a matter of detecting outright deception but rather one of deciding whether we should grant our interlocutor epistemic authority. In detecting deception, we likely use other mechanisms to assess others' intentions, which in turn might influence our reality-monitoring decisions.

2.5.2. VERIDICAL RECOLLECTION AND EPISTEMIC VIGILANCE

Viewing episodic memory as striking a balance between the productive and receptive sides of communication can make sense of the confusing interplay between veridicality and malleability, described in Chapter 3. Similar to reasoning (Mercier & Sperber, 2011), the evolution of episodic memory systems should have been subject to an "arms race" between senders and receivers of communicated information about past events (Jablonka, 2017). Whereas senders have an interest in inducing in their audience a representation of the past that is to their benefit, receivers are interested in acquiring useful (i.e., true) information. Thus, the better senders are at manipulating their audience's beliefs about the past to their own benefit, the better receivers should be at discerning true from misleading information. Both sides of this interaction therefore require the capacity to represent the past accurately.

On the one hand, if episodic memory were never true, it would not convince anyone. Thus, speakers should be sensitive to how informed and skeptical their audience is and

consequently be more careful about what they commit to (i.e., exert more effort in checking their own memory representation for its believability). Receivers, on the other hand, should be sensitive to the interlocutor's intentions and (if available) spend more cognitive resources to monitor the believability of her utterances.

Thus, the epistemic vigilance functions of episodic memory coincide with the epistemic route from memory content to belief: we are able to form and revise beliefs on the basis of episodic recollection because this enables us to guard against others' incompetence and deceptive intent in communicative interaction. This perspective then gives us an explanation for why (and when) we should expect episodic memory to be veridical: epistemic vigilance requires sensitivity to the actual past to enable us to review others' claims and decide when to revise our own beliefs on the basis of such claims. Moreover, the fact that we can expect others to be vigilant, and as such sensitive to the truth, should force us to stick to actual events to the extent that others can monitor us in communicative interaction. Thus, the construction process in episodic memory should be sensitive to the communicative situation we find ourselves in. In cases in which we face a skeptical audience, which raises the costs of being found unreliable, or when we are scrutinizing someone else's claims on the basis of our own memory, construction should aim at accurate event representation.

2.6. A SUMMARY: EPISODIC MEMORY FORMAT EXPLAINED

At this point, it is worth giving a short summary of the view I am proposing. In Chapter 3, I have argued that episodic memories are metarepresentations of event memories: while event memories are about past events, episodic memories are about the source of our representations of those events. In view of these differences, the function of episodic memory

cannot simply consist in the representation of specific past events. Instead, episodic memory allows us to represent why we came to hold (i.e. the source of) a given belief. Its function is thus likely to be found in the domain in which the explicit handling of sources is most important: communicative interaction.

One way in which source information is important in communication is because it can serve as a reason (Mercier & Sperber, 2011; 2017; Mercier, 2016). On the side of the sender, reasons are important insofar as one can supply them to convince an interlocutor who would otherwise not accept what one has to say based on trust alone. Thus, if someone can point to whatever caused them to believe something (their own reasons), this might be good enough for others to believe it, too. On the side of the receiver, one has to be able to tell good reasons from bad ones when deciding what to believe. This also means that the better senders are at giving reasons the better receivers should be at processing and scrutinizing those reasons, and vice versa. Moreover, source information can serve to set the conditions under which we should revise a belief in the face of new information.

Source information is also important in communication because it allows speakers to regulate their conversational commitments. Making an assertion commits the speaker to the truth of whatever she asserts: the speaker accepts that if she is found to be wrong she will incur a direct or reputational cost. The fact that the speaker is willing to incur such costs can serve as a signal for her audience to accept whatever she claims. Thus, the stronger a speaker commits to a claim, the more convincing she should be (Mazzarella et al., 2018; Vullioud et al., 2016). Therefore, speakers should be able to regulate such commitment appropriately since over- (or under-) commitment can be costly. One of the basic ways in which such commitments can be regulated is by claiming or deferring epistemic authority about whatever one asserts (McMyler, 2007). If the speaker claims to have acquired the information in question first-hand,

she at once claims epistemic authority and makes herself directly accountable for the truth of her assertion. This should in turn cause the speaker to be more strongly committed and hence more convincing compared to a case in which she defers accountability to another, second-hand source.

Episodic memory thus allows us to do two things: (1) represent the grounds on which we formed a given belief in the first place, which we can then transmit as reasons to others or use to decide when to change our mind; and (2) regulate the extent of our commitments in discourse by highlighting whether a given event representation originated in our own first-hand experience or not (see also Jablonka, 2017; Poole, 2008; for a view applying a similar idea to collective memory see Seeman, 2016). Of course, this is only the first layer in a complex web of potential source information. On the basis of an episodic representation we can discern whether we have seen, heard, inferred etc. information about a given event (Johnson et al., 1993). Such more fine-grained source distinctions are important because they allow one to answer potential challenges to one's authority ("How do you know?") more precisely than simple expressions of confidence would (Vulioud et al., 2016) and might further be useful in deciding what is informative for one's audience (Nagel, 2015).

Taking a perspective from human communication on episodic memory can therefore illuminate its format in a functional light.

1. *Epistemic generativity* allows us to (meta-)represent the sources for our beliefs about past events so that we can give these as reasons in testimony or use them to decide when to revise our beliefs.

2. *Autonoesis* delineates those of our claims about the past for which we can assert epistemic authority.

Beyond the above features that discriminate episodic memory from event memory, my analysis also accounts for why episodic memory is simultaneously congruency prone (risking to be false) and aiming at veridicality. The fact that event simulation in episodic memory is congruency prone allows us to effectively argue for the beliefs we already hold. Nonetheless, episodic memory is commonly veridical because it serves a role in epistemic vigilance, which requires some degree of sensitivity to actual occurrences.

3. Conclusion

In this chapter, I offered an account of human, mature episodic memory in functional terms. I have explained of the metarepresentational structure of episodic memory in terms of its role in communicative interaction. According to this view, autonoesis allows us to determine when and how to assert epistemic authority in negotiating the past. In effect, episodic memory allows us to communicatively support our interpretations of the past. This view can make sense of a range of empirical evidence: most importantly, why episodic memory construction has the tendency to confirm what we believe about the past and why it is nonetheless commonly veridical.

One consequence of this analysis is that episodic memory should be taken to be human specific. Other accounts arguing for this conclusion have been criticized for being unfalsifiable because they do not offer behavioral markers that could differentiate between autonoetic and non-autonoetic forms of event memory. My account identifies a clear function for autonoetic remembering (the negotiation of epistemic authority), which other animals, in the absence of a communication medium capable of conveying justifications, do not need to fulfill. Thus, from our perspective, it seems unlikely that other animals (and very

young children) would have the capacity for entertaining autonoetic memories, simply because they do not need them.

Another consequence of my account is therefore that the capacity for episodic memory and the capacity to communicate about the past linguistically should be importantly connected both developmentally and constitutively. Although I have not made any specific claims about development, there is at least correlational evidence from developmental psychology suggesting that the capacities for episodic memory and communication about the past are connected (e.g., Nelson & Fivush 2004). Childhood amnesia is generally thought to end between the ages of 3 to 5 (Hayne & Jack 2011), the same time when children begin to be able to use source information productively (Drummey & Newcombe 2002; Gopnik & Graf 1988; Robsinson & Whitcombe, 2003; Whitcombe & Robinson, 2000; Wimmer et al., 1988) and start to display epistemic vigilance (Clément, Koenig, & Harris, 2004; Mascaro & Sperber, 2009). In fact, infants (Bauer & Leventon, 2013) and young children (Burns et al. 2015; Király et al., 2018; Mullally & Maguire, 2014) can recall and make use of event information, suggesting the operation of constructive processes resulting in event memories. However, only after the age of 3 do they become able to use this information as source information in communication (Haigh & Robinson, 2009). Moreover, as predicted by my account, it has been shown that children master the use of evidentials (grammaticalized markers expressing the speaker's source) before they master their comprehension (Ünal & Papafragou, 2016). Finally, in a recent study, we show that children who face disagreement from an interlocutor are more likely to remember their own sources (Mahr et al., 2019). These findings invite further investigations of the relationship between the development of episodic memory and communicative expertise.

More generally, the account offered here is merely a functional one and does not make

precise predictions about the information processing mechanisms involved. The function I propose could be implemented by a range of different mechanisms. Nonetheless, the account predicts that the main achievements in episodic memory development occur as a consequence of the development of retrieval mechanisms. Encoding mechanisms are important for a much wider range of capacities, most of which are not, in fact, connected to our capacity to communicate about the past.

Finally, how can this account of the function of episodic memory explain the functions of our intuitive notion of episodic memory? I want to suggest that our intuitions about 'what remembering is' are essentially structured so as to reflect the fact that remembering ought to confer epistemic authority. As I have argued in Chapter 3, episodic memory is structured so as to produce certain intuitions about its own operation. Most central among those intuitions is the idea that remembering is distinguishable from imagining because it is the outcome of a direct causal link between a past experience and a current mental state. In effect, it is this causal link which confers to the remembering (and not the imagining) subject epistemic authority about what occurred in the past. On my view then, the reason why episodic memory produces erroneous intuitions about its own operation aimed at its own origins in past experience is due to its communicative function: Remembering aims to establish epistemic authority about the past. In effect, from this perspective, the reason why humans have developed this specific concept of remembering in the first place is likely due to the central socio-communicative role of epistemic authority about the past. Why, however, is epistemic authority about the past important enough to warrant the development of a system producing these intuitions? In the next chapter, I will turn to this question.

CHAPTER 5

WITNESSING, REMEMBERING, AND TESTIFYING: WHY THE PAST IS SPECIAL FOR HUMAN BEINGS

Chapter 4 has focused on generating a functional explanation for the structural features of episodic memory. But what arguably is at stake in an explanation of episodic memory function is not only its metarepresentational nature. After all, these are aspects shared with many other aspects of cognition supporting human communication (Mercier & Sperber, 2017). What makes these features interesting is rather the content to which they pertain in episodic memory: representations of specific past events. The question that I have yet to answer is why such a representational structure should be necessary for this content in particular. Why did humans develop a specific mechanism regulating their communicative commitments about past events? Why would we ever want to convince others about a particular version of history, and why do we care what others assert about the details of events they experienced in the past?

1. THE PAST IS SPECIAL

For human beings, the past is special. We think of the past as defining almost all aspects of our identity: where we belong, who our friends are, what our social status is, what kind of person we are. We also love to talk about the past. We share much of our (emotional) experiences with others (Pasupathi, McLean & Weeks, 2009; Rimé et al., 1998) and, according

to one estimate, 40% of our conversational time is spent with telling stories about past events (Eggins & Slade, 2005; Hirst & Echterhoff, 2012). In fact, humans seem to have a 'retrospective bias' in their conversational behavior: we talk about our personal past two to three times as much as about our personal future (Demiray et al., 2018).

The special status of the past is also reflected in the fact that humans seem to have a specific way of mentally representing it. In contrast to other types of knowledge about the world, humans can represent past events via a dedicated 'episodic memory' system. As I have argued in Chapter 4, while other memory systems allow us to simply 'know' what happened in the past, episodic memory lets us 'know how we know' what happened. When we remember a past event, we do not just remember the event, we also know that we experienced it. In other words, episodic memory allows us to become *witnesses* of the past and thus give testimony about it.

The past is so all-pervasively important for us that it seems odd to ask why this might be the case. However, if it is true that episodic memory is a special way of representing information about past events that lets us 'know how we know' about them, then why did we develop such a system? In other words, what is special about past events that requires a special, metacognitive mechanism to think about them? This question, I suggest, is just the question of why past events are especially important for human beings.

My answer to this question will rely on the observation that, for human beings, specific, past events do not only have physical but also social effects that may not leave physical (but only mental) traces behind. For this reason, they require a dedicated capacity allowing us to negotiate them effectively in communication. For some other species, it is also true that certain social relations are modulated by unique events, which requires that individuals

somehow keep score of who did what to whom. I will argue, however, that 'keeping score' only requires the capacity to remember unique events as such to the extent that one wants to communicatively transmit such scores. Because human life depends to an extraordinary extent on the capacity to influence, and be influenced by, others through communicative interactions, it requires a dedicated capacity that allows us to handle events as reasons in making and scrutinizing claims about how commitments and entitlements should be distributed.

I will proceed as follows: first, in Sections 2 and 3, I will tackle the main question of this chapter. What is it about specific, past events that requires a dedicated mechanism allowing us to modulate our conversational commitments in communicating about them? The availability of particular past events in a metarepresentational format facilitates their transmission. But why is the transmission of this information important enough to require such facilitation? In Section 2, I will explore to what extent, the transmission of information about specific past events can serve for others as 'vicarious experience' to learn (i.e., draw inferences to) general knowledge from. I will argue that while there are ways in which claims about specific past events (grounded in remembering) can facilitate the transmission of generic beliefs, there are usually more effective ways of doing so not relying on reference to the past. Not all inferences are such that they are optimally transmitted by reference to specific past events.

In Section 3 I will then argue that the type of inference that is optimally suited to be transmitted by testimony about the past is inference to token effects from token causes. Token causal inferences are particularly important in so far as they establish present social effects like commitments, entitlements, and obligations. Since a large part of our social

ontology is constituted such that it relies on representations of history, testimony will sometimes be helpful in coordinating it.

In Section 4 I will therefore develop an account of how the dependency of many 'social facts' on particular past events might make the communication about these events necessary. On this view, transmitting the events that causally ground a given social fact will sometimes be the only way in which the existence of this fact can ultimately be established. To the extent that there is no other way to independently track the social effects of a given event, testimony about this event will be important in order to signal its existence and thereby coordinate the shared representation of social reality with others.

One way of viewing the project of this chapter is as giving an account about the emergence and social significance of the human ability to bear witness and give testimony. An act of testimony is *an account about the past which is claimed to be based on first-hand experience*. The term 'testimony' is sometimes (especially in philosophy) used to refer to any act of social information transmission. This is emphatically not the way 'testimony' will be used here. Instead, again, 'testimony', as I will use the term, refers to an account about the past based on first-hand experience and the epistemic authority such experience conveys.

2. THE PAST SUPPORTS LEARNING: GENERICS AND REPUTATIONS

If the account of episodic memory I have developed so far is correct, remembering allows humans to give testimony: it allows us to decide when we can speak about the past as witnesses, that is, on the basis of first-hand experience. Episodic memory goes beyond event memory because it has source information 'built into' the event representation. But why are

representations of specific, past events important enough to require the inclusion of source information facilitating their transmission in the first place?

One way to approach this question might be by thinking about how knowledge about the past can be relevant to our fitness. Changes in fitness can inherently only exploits possibilities in the present and future (Klein et al., 2002) but there are two obvious ways in which knowledge about the past can nonetheless be fitness relevant. On the one hand, of course, knowledge about the past might support learning of regularities in our environment. If we know what happened, we might be able to use this information to inductively infer regularities in the way our environment works and therefore form appropriate expectations about what will happen. On the other hand, however, knowing about the past can be important because some past causes have effects that only manifest after some time in the future. In this way, knowing what happened might allow one to predict what will happen or what is the case. Now, can we apply these insights to the question of when the *transmission* of information about the past might be fitness relevant? After all, episodic memory is structured so as to facilitate the transmission of information about the past and the special status of the past seems to be particularly prominent in human social life.

It might seem plausible to answer this question by pointing to the fact that the communicative transmission of a past occurrence could function as quasi-experience for others to form judgments about. If we can transmit our own experience to someone else, to the extent that our interlocutor believes us, she might vicariously learn from this experience just as if it was her own. This fact alone would be a good reason to sometimes require justification for claims about the past. However, not all judgments are equally well transmitted in this way: learning from social information transmission will usually benefit most from generic statements rather than claims about specific events. After all, one of the

greatest benefits of human communication is that we can transmit generic information directly to others, without being reliant on individual learning episodes (Csibra & Gergely, 2011).

Moreover, it is not clear what role the 'pastness' of our experience would play in allowing others to learn from it. The transmission of information about specific *past* events is not identical to the transmission of information about specific events in general. That is, instead of asking what we can vicariously learn from the transmission of information about specific events, we have to ask what we can learn from the retrospective representation of such events that is important enough for event information to play a role in its transmission.

In what follows, I will therefore explore what kinds of judgments (1) are particularly sensitive to the kind of evidence provided by claims about specific past events and (2) carry particularly high social consequence so that humans would care about, and consequently regularly require additional reassurance in their transmission.

2.1. PROJECTIBILITY, INDUCTION, AND GENERIC BELIEFS

When asked what kind of inferences are well supported by reference to past events, most would feel inclined to point to induction. Clearly, to the extent that a given judgment is supported/supportable by inductive inference, it will benefit from reference to past experience. One way to spell out this intuition is to say that testimony is appropriate in facilitating the transmission of a given judgment, to the extent that this judgment is projectible (Goodman, 1983). Roughly, a judgment is projectible if it licenses generalization from a circumscribed sample to a general conclusion, that is, if it licenses induction. The kinds of representations that are sensitive to induction are generic beliefs.

We might thus care about what happened on particular occasions in the past because such events increase the potential sampling base behind our inductive inferences. After all, one way to arrive at a generic belief is by generalizing over specific instances. Thus, the claim that "It rains on Thursdays in Los Angeles" could be supported by pointing out that it rained when I was there on Thursday last week. However, while inductive learning can be supported by evidence from specific past events, neither the specificity nor the pastness of such events is important for such learning (if it is Thursday *today* and it is raining in Los Angeles *now* this takes nothing away from the inductive power of the event). What matters instead is that the specific instance of an occurrence follows a regularity that allows for generalization. Thus, generics are only sub-optimally justified by retrospectively pointing to particulars simply because individual cases might not say much about the general pattern under scrutiny. If one inductively generalizes over a number of instances, one disregards exactly what is particular about each one. The effectiveness of pointing to a specific experience in justifying a generic claim will therefore commonly be limited simply because that experience could be an outlier.[22] The fact that it rained last Thursday in Los Angeles does not necessitate, after all, that it will normally rain there on Thursdays.

Moreover, while pointing to specific past events can be helpful for the justification of inductively derivable generics, generic beliefs are sensitive to all kinds of evidence. Your general meteorological knowledge, for example, might tell you that the weather is unlikely to conform to the days of the week. Thus, generic beliefs are not dependent on reference to specific past events in order to be justifiable and can also be effectively transmitted by reference to other generic facts one holds true (Prasada, 2000).

[22] To be clear, I am not claiming that claims about history cannot be used in order to support inductive generalization; they are simply not well suited to do so.

2.2. BOUNDING AND EXEMPLIFYING GENERICS

There are, however, two other potentially more effective ways in which the transmission of generics can be supported by claims about specific events. First, as Klein et al. (2002; see also Cosmides & Tooby, 2000) have pointed out, claims about specific events can set bounds on how far a generalization might extend. Going back to the example of "It rains on Thursdays in Los Angeles": pointing out that it did *not* rain when I was there last Thursday provides a good counterexample. The universal version of this assertion ("It rains *every* Thursday in Los Angeles") can simply not be true if this specific event occurred. Knowledge about specific events can therefore allow listeners to debate the scope of an assertion. The bounding function of specific events seems particularly useful for the purposes of epistemic vigilance: if we are confronted with a universal claim, but we can come up with a specific instance in which it did not hold, we should, if at all, only accept a more modest version of the claim in question.

Second, instead of being just one more data point for an inductive generalization, communicated information about a specific event might serve as an exemplar (Shafto et al., 2008; 2014). A general pattern might be 'illustrated' and thereby supported by pointing to one specific, diagnostic instance in which it occurred (cf. "strong sampling," Xu & Tenenbaum, 2007). You might, having never encountered a panda bear, wonder whether they are dangerous. You ask a zookeeper, who tells you that he was bitten by one once. On the one hand, as discussed above, this might cause you to inductively increase your belief in the hypothesis that panda bears are indeed dangerous. On the other hand, however, the simple fact that the zookeeper chose this specific episode from his experiences with panda bears to answer your question should cause you to treat this information as being diagnostic of a more general pattern of panda bear behavior. In other words, this episode would not make the

hypothesis that panda bears are dangerous more believable because it would provide one more instance to inductively generalize from. Instead, the listener will assume that the speaker picked that episode to share because it provides the best example to learn from and this in turn would make the target claim more convincing. A specific event can therefore serve to justify a general claim in virtue of its exemplary character.[23]

One benefit of pointing to specific events as exemplars is thereby that one does not have to make explicit the target claim one aims to transmit. Simply pointing out that "I was bitten by a panda bear once" will sometimes be enough to make one's audience infer that they must be dangerous. Providing exemplars, however, will likely be mostly necessary when reasons in support of a prior claim are requested. After all, as mentioned above, one of the main advantages of communication in the first place is that we can transmit ready-made generalizations to others. Only when challenged, will pointing to a specific episode (in the form of testimony) become necessary.

According to what we have discussed so far then, we should expect people to care about what happened at specific occasions in the past primarily because (1) past events can set bounds on generics, allowing us to evaluate and contradict them, and (2) because past events can serve as exemplars for transmitting generics to others in argumentation and teaching.

[23] The process by which listeners arrive at the assumption that a speaker in a given context is making an 'exemplar claim' can be described in two complementary ways. On the one hand, the listeners' assumption can be viewed as a 'sampling assumption' as described in Bayesian models of teaching (Shafto et al., 2008). Here, the listener takes the speaker to have picked this event from a distribution of other statements/events that would support the same claim because it optimally exemplifies it. On the other hand, the assumption that a speaker is providing an exemplar can be described as the outcome of a pragmatic inference process which relies on an 'assumption of optimal relevance' (Sperber & Wilson, 1995): In order to arrive at a speaker's intended meaning, listeners assume that whatever the speaker is stating 'maximizes' inferential pay-off for her listener.

2.3. DISSEMINATING REPUTATIONS

Humans can teach, argue for and evaluate almost anything by pointing to exemplifying events (or chains of events). In these cases, providing reasons in the form of specific events functions according to the same principles as argumentation in general (Mercier & Sperber, 2017). A claim will require justification if the audience does not trust the speaker enough to accept her claim on that basis alone. Consequently, justification will be required when the stakes of being mislead or the incentives to mislead are high.

While this is true in many specific contexts, a domain where these conditions are met very consistently are claims affecting others' reputation. According to Dunbar (2004), the most common topic of conversation are social evaluations. As much as 65% of casual conversation concerns social topics (about others' interactions, behaviors and traits) (Dunbar, Duncan & Marriott, 1997). People's interest in others' behaviors and interactions is enormous even when they themselves are not involved (DeScioli & Kurzban, 2009), and this pattern does not seem to be exclusive to Western societies: Zinacantan people in Mexico similarly have been reported to spend 78% of conversational time talking about such social topics (Haviland, 1977).

This phenomenon is commonly termed 'gossip' (Foster, 2004) and has been proposed to be essential in the stabilization of group living (Dunbar, 1998; Wu, Balliet & Van Lange, 2015, 2016a, 2016b). Dunbar (2004) has argued that the transmission of social evaluations plays an essential role in the stabilization of our conditions of communal living. The reason for this is that information about someone's past behaviors are often taken to be diagnostic about her future behavior; that is, that such information potentially licenses trait inferences.[24] Once trait

[24] Crucially, such trait-inferences do not have to be valid as research on phenomena such as the 'fundamental attribution error' (Ross, 1977) and the 'correspondence bias' (Gilbert & Malone, 1995) shows.

judgments become shared across a group they develop into reputational information. By disseminating reputational information through a given social group, the transmission of social evaluations allows us to go beyond our personal experience when assessing the state of our social network and the dispositions of others (Sommerfeld et al., 2007). This in turn is taken to fulfill 'policing' functions (Foster, 2004), effectively implementing a form of social control because what someone believes about others' traits will determine who she associates with and how she interacts with them. Therefore, if one can manipulate others' opinions about a specific person, one can effectively control coalitional associations as well as cooperative opportunities. This explains why we gossip so much about others' past behaviors and are careful about tracking (our own and others') sources of this information (Wilson et al., 2000). We have to justify and be vigilant against being mislead about claims affecting other's reputation because of the various ways one could take advantage of changing someone's reputation (Hess & Hage, 2006). Indeed, one's conversational commitments are rarely more important than in the domain of gossip. Where a piece of gossip originates from, and how far a given speaker is removed from having experienced the episode in question, are crucial both for how believable the gossip is and who is responsible for it (Giardini & Conte, 2011).

3. THE PAST GENERATES ENTITLEMENTS, COMMITMENTS, AND OBLIGATIONS

3.1. TOKEN CAUSES AND THEIR SOCIAL EFFECTS

So far, I have identified generic beliefs as one kind of judgment the transmission of which can be supported by referring to specific past events. Moreover, I have argued that the transmission of judgments about others' traits and dispositions are a domain where

incentives to mislead and risks to be misled are regularly high. That is, we should expect source claims and the modification of conversational commitments they allow (i.e., testimony) to be particularly important in the transmission of trait judgments; i.e., in gossip. Claims about specific past events are, however, not only relevant in the transmission of generic beliefs, and generic beliefs can be transmitted without ever referring to such events. In fact, as mentioned above, past events play a role in the transmission of generic beliefs not necessarily in virtue of the specificity or pastness of these events but rather as input to inductive learning machinery or as examples of a more general pattern. To explain why the past *for its own sake* seems to have such a special status for human beings, we should look for a domain in which retrospective reference to particulars is *required*.

Further, the transmission of generics, including reputation, may not provide a sufficient evolutionary pressure for the development of testimony because the main beneficiaries of teaching and of the spread of reputational information are the recipients of such communication: they acquire knowledge to be used in the future, which may have fitness consequences. While there might be other factors that would make these types of communication fitness enhancing for communicators (e.g., reputational gain), these would only be indirect benefits. It is therefore worth asking if there is an explanation that relies on direct fitness enhancement for the person who gives testimony. How could the speaker's, as opposed to others', fitness benefit from testimony?

3.2. TYPE AND TOKEN CAUSES

One domain where this might be the case is causal judgments. The relationship between past and present is commonly conceived of in terms of causal relations. People constantly

infer causal relations between events unfolding around them (Gopnik, 2000). Nevertheless, while causes are events, the representation of causal relations as such does not require the representation of specific events: causes are often represented in terms of 'type causation' ("being shot kills people"). What does require the representation of a specific event, however, is the retrospective inference from a specific, token effect to its token cause ("Mark died because he was shot"). Crucially, such a retrospective inference requires not just the representation of an event as the cause of an effect, but also its representation as having occurred temporally before the effect, i.e., in the past. Moreover, while in principle an unbounded set of causes underlie any given effect, humans commonly represent causes and their effects as standing in a one-to-one relationship; in other words, specific token effects are often represented as having specific token causes. Token causal judgments then have all the qualities that would seem to make the representation of past events necessary in order to link causes to effects.

Token causal judgment is a domain in which particulars (i.e., specific past events and their counterfactual derivatives) are crucial (Campbell, 1996). Epistemic authority about the actual occurrence of specific past events might thus matter particularly in the transmission of causal judgments as causal explanations.[25] Note, that *type* causal explanation requires claims about the past experience of a specific event only in so far as they are relevant to the transmission of generic causal beliefs more generally. Claims to personal experience will, thus, be particularly powerful in the transmission of *token* causal judgments.

[25] In order to play a role in causal judgment itself, the representation of event information is sufficient. Epistemic authority about the event in question (and hence remembering proper) only becomes relevant in the transmission of such causal judgments.

3.3. PHYSICAL AND SOCIAL EFFECTS

The benefits of making token causal judgments may depend on the nature of the effect in question. Finding a token cause for a *physical* effect is an inference to a specific, past event, but establishing this causal relation affects our future fitness only to the extent that it allows us to inductively generalize it (and use this generalization, for example, in planning future actions). While causal thinking is a powerful learning engine allowing us to understand, predict and explain contingencies in our environment, these benefits only partially apply to thinking in terms of token causes: particular instances of causation serve as learning opportunities mostly in light of our capacity for building causal maps from representations of type causal relations (e.g. Gopnik et al., 2004). Moreover, token instances of causation can usually only be interpreted through the application of type causal assumptions and serve as opportunities for learning only in so far as they inform these assumptions. In the physical domain, token causal judgments are therefore most important as inputs for inductive learning mechanisms. As we have argued above, however, inductive inferences are not optimally transmitted through testimony.

This is similarly true for retrospective causal inference: say you arrive in your office one day to find that your computer screen is lying on the floor and does not work anymore. When you ask your office-mate what happened, she informs you that one of your co-workers threw your screen on the ground in a fit of frustration. Regarding the purely physical cause-effect relations at play here, this information will be relevant to you in so far as it informs you that computer screens tend to stop working when thrown on the ground. That is, you will benefit from knowing the physical cause of why your screen stopped working in so far as you can infer a type causal relationship from this specific instance. You might then, for example, fix your screen to your desk so as to avoid it falling or being thrown on the ground

in the future. In this way (i.e., via a type causal inference), information about token causal relations can impact your future behavior adaptively.

For humans, however, causes instantiated in specific past events are often more than opportunities for learning about our physical environment via type causal inferences; they may have important social implications, too. Many physical or biological causes produce not only physical but also social effects. If I manually create an artifact, I may earn rights to use it or own it; if my aunt dies, I may inherit some of her property; if your dog kills my lamb, you may have to compensate me; if I father a child, I may have to contribute to her upbringing; if a landslide destroys my crop, I may be relieved from the duty of contributing to the common good; etc.

Note that while some, but not all, of these events are actions of social agents, all of them produce lasting changes in the physical environment. However, crucially, they also produce social facts: someone becomes/ceases to be the owner of a resource, a father, a debtor, etc., thereby producing entitlements and obligations that have fitness consequences for the future. And because, unlike the physical/biological effects of the same causes, these social facts are not perceivable, only the causes that produced them in the past prove that they obtain. In fact, these causes are thought to play a constitutive role in them. To a large extent, this explains why we are bound to care so much about specific past events (including their actual details).

Going back to the example of the broken screen above: the crucial inference you will likely draw from the information that your co-worker threw your screen on the ground will in fact not be a type-causal relation. Instead, you will form an accountability judgment that serves as the basis for a claim to an entitlement for compensation. That is, in this case, knowing the

specific token cause will have an effect on your future fitness not (primarily) by allowing you to learn about type causal relations but by allowing you to infer and transmit the social effects produced contingently with the physical ones.

3.4. When testimony is necessary: Communicatively generated commitments

People see certain physical events as generating (and sometimes even constituting) social effects. Thus, occurrences of certain specific past events can inform us about present and future social entitlements and obligations, and since these social effects exist primarily as mental representations, testimony about such events can be an important argument during negotiations of entitlements. However, strictly speaking, testimony is not the only way to prove the occurrence of such events. Even though the social consequences of such events (e.g., the entitlement for compensation) are not perceivable, their contingent physical effects (such as the broken screen in the example above) can still be traceable. This in turn, at least in principle, may allow retrospective inference from effects to their causes without relying on the testimony of others. One can always try to do the detective work backwards from the physical effects to infer, and argue for, the cause and thereby for its social effect. The craftwork on an artifact may show who created it; the exhumed corpse of my aunt can prove that she really died; the injuries of my lamb may reveal that your dog was the culprit; fatherhood can be inferred from DNA tests or from facial resemblance; the change of the landscape provides evidence of a landslide; etc. Thus, while testimony (and the episodic memory it requires) is useful to argue for the validity of a given social fact in all of these cases, it is not mandatory: Contingent physical effects may allow us to infer the past physical (or biological) causes that induced the present social facts in question.

Nonetheless, once the ability to represent the social effects of events emerged in human evolution, it likely made the ability to refer to the past on the basis of remembering it (i.e., testimony) extremely useful. Once in place, moreover, this ability could then have given rise to new forms of commitments that do not necessarily rely on traceable physical effects: promises, agreements, bets, and marriages are all examples of social effects which do not necessarily leave physical traces behind. Instead, they are generated by communicative acts. These instances of communication normally have no correlated, lasting physical effects. Therefore, not only their social effect but also the cause itself exists only in the mind of the participants.

If Margaret promises Elena that she will be back home by 7pm, the effect of this promise (i.e., that Margaret is now committed to a certain behavior) is not observable; it survives – if at all – only in the minds of Margaret and Elena (and any possible witnesses). Nonetheless, the promise-commitment relation here seems to be of the same kind as the token cause-token effect relation described above.

The proof that such a cause occurred could only come from testimony - hence the necessity of episodic memory. In a sense then, social effects of this sort have an inherent 'dual temporality'. On the one hand, they are about the future: a promise obliges to a behavior, an ownership transfer entitles the beneficiary to privileged use, etc. Nonetheless, once established, the existence of the ensuing obligations and entitlements can be justified or proven only by reference to the past event that established them. Without the ability to communicatively refer to the past, such practices could not have developed.

Crucially, it is exactly because of the dependence on testimony of these causal events that societies developed ways to ensure their provability by recruiting witnesses for ceremonies,

and (only more recently) by creating correlated physical effects of these 'non-physical' causes in the form of documents (contracts, certificates, memoranda, bills, records, etc.; e.g., Basu et al. 2009). That is, the ephemeral nature of the cause-effect relationship in social commitments induced the cultural evolution of a host of 'commitment devices' (Fessler and Quintelier 2013), designed to alleviate reliance on individual memory alone by requiring the commitment event to become physically traceable in one form or another. Further, events grounding explicit commitments are often ritually structured so as to be public and easily referable: a promise is accompanied by a handshake, a marriage by a ceremony, etc. Making a commitment public, for example, not only increases the cost of possible defection but also coordinates the representation of this social fact in the community. It is worth noting, however, that while unperceivable social facts (ownership, social status, etc.) are frequently signaled publicly to make sure that others are aware of them without having to prove them again and again, these documents do not simply indicate that certain social facts obtain but are also designed to prove that the specific cause that brought them about indeed occurred (this is why date and place, which together individuate a specific episodic event, are included in them).

Before I continue, let me recap the argument I have developed in the last two sections. Information about specific, past events can be used for various purposes. It can support inductive inferences about projectible properties of objects, agents, situations, and causal relations, which support the acquisition of generic knowledge about kinds, individuals, and type causal relations. However, this purpose can be achieved in various other ways as well, and so it does not require the preservation, representation, or testimony about, past events. A special subset of token causal events, however, produces not only (or no traceable) physical effects but also social facts that allocate entitlements and obligations to specific

individuals or groups. These social facts are generated by their own token causes, and therefore the ultimate proof of their existence is evidence of the occurrence of these token events (see also Poole, 2008). Episodic memory and testimony of past events can thus be crucial for the stable maintenance of such social facts in the community (see below). In fact, reliance on communicatively established commitments (such as promises), which may not leave any physical trace behind, could not even emerge without cognitive mechanisms that ground both prospective memory, to ensure fulfillment, and retrospective memory and testimony, to ensure accountability.

4. THE HISTORICITY OF (MANY) SOCIAL FACTS

What is then the relationship between social facts (obligations, entitlements, commitments, etc.) and episodic memory? It has been proposed that memory capacities are necessary for enabling certain forms of social interaction (e.g., Stevens & Hauser, 2004). However, tracking social relations can also be accomplished by cognitive 'bookkeeping' mechanisms that keep and update scores of interacting agents upon each encounter. A given interaction would then be interpreted depending on the score of each agent involved (e.g., Nowak and Sigmund, 1998). In this way, nothing about the event in question has to be remembered, because its outcome simply updates such a score. Say, Isa lends 5 Euros to Rahmeed. In order for Rahmeed to reciprocate and pay Isa back, all he has to keep in memory is that he now owes Isa 5 Euros. Nothing else about the lending event itself has to be remembered.

A number of different authors have proposed that such a tracking mechanism could have been implemented through an 'attitudinal' (Brosnan and DeWaal 2002) or emotional scoring system in non-human animals (Schino and Aureli 2009; see also Gervais and Fessler 2016). It

is likely that in many situations exchange-related information is tracked in a similar manner in humans (Bell et al. 2017; for a modeling approach related to this issue see Kleiman-Weiner et al. 2016). The representation of specific, past events is therefore not a requirement for maintaining stable pairwise social relations. Learning from 'exchange events' is similar to learning from events which have no social consequences: one can draw inferences from such events without storing much of what happened.

While commitments and entitlements can only ultimately be proven by reference to specific past events, the representation of their existence does not depend on the capacity to recall specific past events. In order to know that John owns his car, you do not have to remember anything about the event in which he acquired it (even though you have to assume that there was such an event). Why should the transmission of the privately represented social effects of a given event be important then?

4.1. MAINTAINING AND STABILIZING SOCIAL FACTS

The social effects produced by ordinary physical or biological events have to be maintained by some forms of public representation, such as face-to-face communication. After all, they often exist only in the minds of individuals, and communication is the main means by which these effects become and remain shared. If they are not shared, social facts do not fulfill their function, so it is in everyone's interest to coordinate them well.

Put differently, in order to become social facts, the privately represented social effects of events have to be shared and agreement about them has to be established. Ownership, social structure, and social roles are good examples here: they may also be marked by permanent public signals to ensure common acceptance even in the absence of direct verbal

communication. Social facts such as these may be generated by token causes, but their shared maintenance depends on communication, and if their existence is disputed, they can be negotiated by reference to the events that brought them about. Social facts inherently depend on public agreement, and to achieve such agreement, the past events grounding a given fact have to be available. This is important not only in cases of conflict. Rather, it is simply not possible to decide privately whether a given social fact indeed applies. While one can represent a social fact as such without entertaining the (historical) reasons why it obtains, in communication such reasons might have to be explicitly invoked as the ultimate argument for its existence.

Crucially, for humans, an event can be ambiguous as to what social implications it establishes. A given episode is often important not just because of factual occurrences but for the myriad ways in which these events *could have* turned out. What a person *did not* do, and what her intentions were in acting, for example, are essential in computing the ways in which commitments should be distributed. While humans have a host of specialized cognitive mechanisms that enables them to carry out such computations online, the transmission of the conclusion will often require justification. Distinguishing between, for example, incompetence and malevolence will sometimes require that one refers to details of the specific action in question. While malevolence should trigger punishment or ostracism, incompetence does not necessarily call for these reactions (see Nowak & Sigmund, 2005 for why the ability to make such a distinction might be important). The social coordination of the representations of implications of specific events will thus often unavoidably require communication about such events (for a similar point see Pietraszewski, 2016).

Testimony, therefore, helps to maintain and coordinate the validity of entitlements, obligations, and commitments within a social group. While it might also serve other

important functions, testimony can play a 'signaling' role in advertising the existence of certain social facts: I claim that this knife is mine and I justify it by the fact that I made it, or that I inherited it, or that it was donated to me, etc. It does not necessarily require an open challenge or violation of property rights to make these assertions; in the absence of a permanent symbol system to mark ownership and other entitlements, repeated declarations of social facts may be necessary to maintain their shared nature and to let newcomers know about them.

Note that this conclusion does not require that communicatively coordinating social facts should always, or even necessarily, involve pointing to specific past events. After all, beliefs can be transmitted on the basis of trust alone, without requiring the representation or transmission of reasons. Moreover, social effects affecting a whole community are commonly structured so as to be purposefully independent of individual testimony. Events establishing important social effects are ritualized or designed so as to generate public knowledge from the outset by either generating concomitant physical effects (e.g., documents) or many witnesses. In this case, testimony becomes less important.[26] In fact, the motivation to make such events independent of individual testimony has likely lead to the development of technologies allowing for the generation of public knowledge.

Once public knowledge about a given social fact has been established it will rarely be challenged: marriages or kinship relations rarely, if ever, become a matter of dispute.

[26] Events that only affect a subset of group members are therefore often more likely to become the subject of transmission by testimony because these are often not structured so as to produce public knowledge or leave intentional records. This is another reason why testimony about other group members is common in gossip: this information is not just interesting because it potentially allows interlocutors to draw trait inferences but also because of the fact that we can effectively transmit commitments, entitlements, accountabilities etc. in this way. The fact that the mayor has been cheating on his wife with his secretary, for example, will likely become the subject of testimony of individual community members while the fact that he is the mayor (while similarly dependent on a specific past event) will rarely have to be testified to because commonly everyone already knows about it and there are documents proving it.

Therefore, social effects that affect the whole community will often not depend on testimony to be maintained because they generate public knowledge or are otherwise made traceable for everyone. One reason for this, however, is that they are assumed to be ultimately appropriately justified by past events, which can be made available one form or another in case of doubt. Only the possibility of pointing to the establishing event in any given case ensures that disagreements about the fact in question could be resolved in principle. Many social facts are simply such that only their establishing event can ultimately arbiter whether (or in what way) they obtain. If there was no way to refer to or to make available these events, there would be no way to ultimately ensure the appropriateness of claims about such social facts, and consequently to justify the ensuing entitlements and to enforce their fulfillment.

If my analysis here is correct then the capacity for testimony, underpinned by human episodic memory, must have enabled the capacity for coordinating certain social facts in the first place, and for generating new types of social facts that could not even exist without testimony. The reason for this is *not* that such commitments require episodic memory to be cognitively traceable and behaviorally implementable by the individuals involved. Rather, they require testimony to be shared in a community. This is because in some cases there could be no fact of the matter whether a given obligation, entitlement or commitment applies without the potential of testimony about the specific, past event causally grounding the social effect under dispute.

4.2. MOTIVATED REMEMBERING, MEMORY BIAS, AND NARRATIVITY

The above argument predicts that episodic memory is motivated (and hence to some extent biased) by design so as to justify one's own present entitlements (see also Lambek 1996). In Chapters 3 and 4 I argued that one way in which such a bias manifests is through 'recollective my-side bias.' Episodic memory construction is more likely to confirm and support our prior beliefs than to contradict them. However, this might not be the only source of architectural bias in episodic memory. As also mentioned in Chapter 3, episodes are neither retrieved nor communicated as atomized particles but as narratives. Testimony is not just given as a series of propositions but it is narrativized in a way that makes it more likely for the audience to draw certain inferences over others. According to Keven (2016), episodic memory retrieval includes a mechanism (referred to as 'narrative binding') connecting isolated event representations by inferring not only temporal ('X happened before/after Y') but also causal ('X occurred because of Y') and teleological ('X occurred so as to bring about Y') relations between them. On this view, episodic memories allow us to understand the past so as to make sense in light of causal and teleological relations between different events as well as their connection to the present (Bietti, et al., 2018).

How did I get to work this morning? I went to the bus stop to get the bus at 7:45 but the stop was closed because of a construction site and so I had to take the metro to work instead. Already this minimalistic account of the events of this morning includes a significant amount of selection and interpretation in so far as certain events and their causal/teleological relations are highlighted and others left out. In order to effectively argue for the validity of a given social fact, making the temporal relations between events available is not enough. Instead, we have to be able to bind events in a way that highlights the causal connections in question (e.g., having been forced to change the mode of transport this

morning added delay to my travel and caused me to be late for work). Narrative binding processes therefore always include a modicum of interpretation: relating events causally and teleologically includes a selection process in which certain events are highlighted over others. Narratives are often effective because they display events as being (causally and teleologically) related in a way that suggests certain conclusions over others. The fact that episodic memory is narrativized follows from its crucial communicative role in establishing social facts in the present through reference to history (for a somewhat related account relating narrativity to the necessity of providing 'excuses', see Breithaupt, 2011).

4.3. ACCOUNTABILITY JUDGMENTS AND THE ROLE OF SOCIAL NORMS

One might propose that testimony not only plays a role in signaling and coordinating the validity of social facts, as I argued above, but is moreover of crucial importance in the enforcement of social norms themselves. After all, the context in which testimony seems to matter most in contemporary societies is the legal domain. Testimony has been investigated by cognitive psychologists mostly in the form of eyewitness testimony for crimes (e.g., Neisser 1981; Wright et al., 2009), and ethnographic accounts have often particularly focused on the role of witnesses in the legal domain (e.g., Gluckman, 1955). Thinking about the relationship between our capacity for testimony and norm enforcement, one might thus conclude that testimony enables the enforcement of social norms by informing others about the violation of those norms. Without the ability to share information about such violations, people would always be dependent on first-hand experience in judging whether a norm has been violated, which would not make it possible for norms to be widely enforced by third parties/communities in general.

In my view, however, the role that testimony plays in norm enforcement is just a special case of the more general role I have outlined above. In essence, sharing event information pertinent to norm violations aims to transmit a judgment - an *accountability* judgment - to establish a social fact about such accountability. That is, while accountability judgments as such are private, they can be justified and thereby transmitted to others by pointing to the event in which a norm was violated.[27] The transmission of such a judgment aims to establish a shared representation of accountability, through which it would become a social fact. The enforcement of the norm in question, however, may follow from the accountability judgment itself, not from its transmission. Only once accountability has been established and is shared within a group, norm enforcement may ensue. The transmission of the norm violation event only serves to coordinate the representation of accountability and is not directly involved in the enforcement of the norm.

Thus, testimony is common in the social negotiation of accountability judgments because additional reassurance in communication about the past is required when the stakes are high. Arguably, as the domain of norm enforcement has become institutionalized, the forms in which testimony is given in this context (e.g., as 'eye witness testimony' under oath, etc.) have become cultural institutions, too. Norm enforcement is facilitated through testimony because the fact of a norm being violated is sometimes not physically traceable. In fact, contrary to other domains, accountability transmission can hardly be alleviated from its dependence on testimony through cultural or technological solutions in principle: norm violations are not usually public and rarely leave documents behind.

[27] Again, while these events will commonly be actions, they do not have to be (the absence of action, for example, can just as well lead to accountability).

Testimony, therefore, will often help in justifying and determining punishment. In essence, however, norm enforcement and our capacity for testimony are not dependent on one another. Norms can be enforced without the involvement of testimony, and testimony is effective and occurs outside the domain of norm violations and their enforcement. It is important to note here, however, that, the role of testimony in the transmission and coordination of social facts, depends, to some extent, on social norms in the first place. After all, what social consequences follow from a given event is commonly governed by social norms. It follows from this point that our propensity to represent (and observe) social norms must have existed prior to the emergence of the role of testimony outlined here.

Nonetheless, while testimony might not play a role in the enforcement of social norms and commitments, it certainly makes them more effective. Our capacity to bear testimony changes the dynamics of social interactions in crucial ways. Testimony transforms the pay-off structure of two-person interactions into one in which third-parties are always at least potentially present. In fact, (the possibility of) report has been shown to be highly effective in promoting cooperation (e.g., Wu, Balliet, & Van Lange, 2015; 2016b).

People intensely care about whether their behavior is being witnessed by others and this concern emerges relatively early in development: by the age of five, children have developed a robust sense of the consequences of someone else witnessing their norm violations (and the communicative forms of aggression that can ensue) and adjust their behaviors accordingly. Five-year-olds behave more prosocially in the presence of peers (Engelmann, Herrmann, & Tomasello 2012), when they believe to be watched (Piazza, Bering, & Ingram 2011), or when their actions are witnessed by others (Leimgruber, Shaw, Santos, & Olson 2012). On the one hand, these effects are likely to be due to the fact that witnessing someone's behavior might cause third-parties to draw inferences about their

traits and accountabilities. On the other hand, people are likely aware that witnesses could pass on their evaluations via testimony. This would spread their judgments to the community, which might in turn motivate alliance recruitment against the observed individuals (Pietraszewski, 2016; Boehm, 2012) and potentially influence their reputation. In fact, gossip is likely to play an important role in this process: while (as discussed above) gossip regulates the spread of reputational information, it also serves as a norm enforcing device by transmitting others' accountabilities through reference to norm violating behaviors.

5. CONCLUSION

The main aim of this chapter has been to give an account of the role of representations of the past that makes it intelligible why they have such a special status for humans. The past plays a particularly crucial role in human social life. For humans, events do not only have effects in their physical but also in their social environment. The representation of such effects impacts our future fitness to the extent that we can establish the validity of the ensuing social effects with others. Commonly, the only way this can be achieved is by retrospectively pointing to the event that produced the social effect in the first place. For this reason, the past becomes important to us and is also frequently contested. It is this circumstance that makes the ability to remember the past in episodic memory particularly beneficial.

Episodic memory allows one to become a witness of the past and give testimony about it. Testimony is a way to facilitate the transmission of information about past events by conferring epistemic authority and increasing speaker commitment. Such facilitation is

required in cases in which one's audience requires additional reassurance about whatever one is asserting. Given the extraordinary importance of representations about the past in deciding what social realities apply, we should expect claims about history to be a context in which listeners regularly require such additional reassurance.

Testimony will thereby be most important in the negotiation and transmission of our own and others' commitments, entitlements, and accountabilities. Communicatively pointing to the past allows us to justify assertions about the existence of social facts, and a large range of cultural practices has developed exactly to alleviate the reliance of social reality on individual memory and testimony.

This view has consequences for how the evolution of episodic memory (the cognitive basis of testimony) must have looked like. Episodic memory might have developed only once humans were able to represent the social effects of the events in their environment. This ability, however, must have required the prior emergence of social norms determining these social cause-effect relationships. After all, if no one represented or followed social norms, the past would lose its importance as a way of coordinating social realities. Once in place, our ability to testify about past events could then have also be used to transmit generic beliefs to others and thereby make reputation dissemination more effective.

CHAPTER 6

GENERAL CONCLUSION

The goal of this thesis has been to give an account of the nature and function of remembering. In order to achieve this goal, I have proceeded in multiple steps. First, in Chapter 1, I argued that the causal theory of remembering is the best explication of our intuitive notion of remembering. I then asked, why this intuitive notion has the specific structure it has: why does our everyday notion of remembering include the idea that our current representation of the past must have been caused by our past experience of it? To answer this question, I argued, we need an account of the cognitive architecture of the episodic memory system.

Therefore, in Chapter 2, I started to give such an account by investigating the structure of episodic simulation, the ability to simulate mental imagery about events. According to this view, episodic simulation consists in a mechanism generating mental imagery supplemented by a propositional, metarepresentational scope syntax specifying what these simulations should be taken to be 'true of'. Through the application of such scope operators either at the input or at the output stage of episodic simulation, this system can output a large variety of event representations differing along four 'dimensions': specificity, temporal orientation, subjectivity and factuality.

In Chapter 3, I then proceeded to argue that the contents of episodic memory correspond to one particular type of output of episodic simulation: specific, past, actual, personal event representations. Such outputs, however, are not yet episodic memories. Instead, episodic

memory corresponds to a metarepresentational redescription of such contents as 'having been experienced'. That is, episodic memory is the result of the ascription of a distinctive propositional attitude ('remembering') to episodic simulations of specific, past, actual, personal, events. As a result, in episodic memory we represent the outputs of episodic simulation as the source (or past cause) of a current belief. Thus, episodic memory goes beyond the outputs of episodic simulation in two ways: (1) by being 'autonoetic' because it includes a representation of its own origin in first-hand experience (and consequently its own representational character, too) and (2) by being 'epistemically generative' because this representational character plays a role in the beliefs we form on its basis: because we take our current event representation to have been caused by past first-hand experience, we believe that the event occurred. The intuition that remembering requires a causal connection between a current representation of an event and a past experience of this event is therefore an outcome of the fact that episodic memory presents itself as having been caused by a past experience. In fact, it is this piece of metarepresentational content that distinguishes episodic memory from other outputs of the episodic simulation system including representations of specific, past, personal events (which I have called 'event memories').

Why, however, is it the case that episodic memory includes a description of its own causal history in past experience? In Chapter 4, I aimed to give a functional explanation for this metarepresentational structure of episodic memory. I argued that episodic memory allows us to represent the source of our current beliefs and that this capacity is particularly important in communicative interaction. On the one hand, the autonoetic character of episodic memory allows us to distinguish those event representations which we formed on the basis of personal experience from event representations acquired in other ways. This is

important in so far personal experience confers epistemic authority in transmitting those representations to others. The reason why claims to personal experience might confer authority and make whatever one asserts more convincing, moreover, should be that it commits the speaker more strongly to the truth of her assertion than other source claims. On the other hand, episodic memory allows us to effectively assess when to re-evaluate a given belief on the basis of new information by making available the grounds on which we formed this belief in the first place. This is particularly important in the domain of communication where we have to continuously re-assess our interlocutor's competence and benevolence.

While this account makes it intelligible why episodic memory has its particular metarepresentational structure, it leaves open the question why this structure specifically pertains to representations of specific, past events. In particular, it remains unclear why the negotiation of epistemic authority about past events should be of such high importance to humans to warrant the development of a dedicated cognitive mechanism. In Chapter 5, I gave an answer to this question by investigating why the past seems to have such a special status in human social life. I argued that, while information about past events can be important in assessing and transmitting inductive inferences (most importantly about other's traits), the special role of the past as such is likely grounded in the way humans think about the social token cause and effect relationships obtaining in their environment. From this perspective, events can have physical (i.e. non-social) effects and social effects. While knowing the cause of a physical effect is important to us in so far as it allows us to draw inductive inferences about the causal relationships obtaining in our environment, knowing the cause for a given social effect is important because it is often the only thing that can prove that this social effect indeed obtains. In contrast to physical effects, social effects are

usually not independently traceable since they mainly exist as mental representations. Being able to refer to a past cause of (say) a present entitlement is therefore often the only way to 'prove' that this social effect indeed obtains. Because it can establish or contest a large variety of social realities, the past therefore has such extraordinary importance to humans. Consequently, epistemic authority about the past can confer important privileges that go beyond the epistemic domain, which is why tracking such authority should be highly adaptive. Therefore, the intuition that remembering should be an outcome of a direct causal connection between past experience and current representation defines our understanding of remembering because it tracks epistemic authority about the past and the tracking of such authority is important as a way to coordinate social realities in the present.

The result of this thesis is therefore a comprehensive account of the structure and role of remembering in cognition, communication, and culture. Of course, I did not and could not explore all of the many possible implications this account has for the understanding of remembering in other domains. There is, for example, much to be said about the interaction between remembering in individuals and groups. The notion of remembering and the special status of the past does not, after all, only apply to individuals but also to human beings in so far as they understand themselves to be a group (Hirst, Yamashiro, & Coman, 2018; Olick, 1998; Poole, 2008). There has been intense interest in the study of human memory and the cultural uses of recollection in the social sciences (a "memory boom," Berliner, 2005; Winter 2001). From the perspective presented in this thesis, it is, in fact, not surprising that remembering should be of central interest to social scientists. After all, if I am right, episodic memory in some sense enables the commitments and entitlements that make up the web of social relationships we are embedded in both as individuals and as members of social collectives. Indeed, the same kinds of justificatory practices that are used in the negotiation

of interpersonal commitments emerge on the collective level in how past events and their commemoration are used in the political arena in the negotiation of collective commitments and entitlements (e.g., Olick & Levy 1997; Pool 2008; Weiss 1997). I therefore take my account to contribute to the integration of these different perspectives on human memory and its uses. Remembering, far from being the intimately private affair we intuitively take it to be, has a fundamentally social dimension.

REFERENCES

Addis, D.R. (2018). Are episodic memories special? On the sameness of remembered and imagined event simulations. *Journal of the Royal Society of New Zealand, 48*(2-3), 64-88.

Addis, D. R., Cheng, T., P. Roberts, R., & Schacter, D. L. (2011). Hippocampal contributions to the episodic simulation of specific and general future events. *Hippocampus*, *21*(10), 1045-1052.

Addis, D. R., Pan, L., Vu, M. A., Laiser, N., & Schacter, D. L. (2009). Constructive episodic simulation of the future and the past: Distinct subsystems of a core brain network mediate imagining and remembering. *Neuropsychologia*, *47*(11), 2222-2238.

Addis, D. R., & Schacter, D. (2012). The hippocampus and imagining the future: where do we stand? *Frontiers in human neuroscience*, *5*, 173.

Addis, D. R., & Schacter, D. L. (2008). Constructive episodic simulation: temporal distance and detail of past and future events modulate hippocampal engagement. *Hippocampus*, *18*(2), 227-237.

Addis, D. R., Wong, A. T., & Schacter, D. L. (2007). Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration. *Neuropsychologia*, *45*(7), 1363-1377.

Aikhenvald, A. I. (2004). *Evidentiality*. Oxford University Press.

Allan, K., Midjord, P., Martin, D. & Gabbert, F. (2012). Memory conformity and the perceived accuracy of self versus other. *Memory & Cognition*, 40, 280-286.

Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin*, *93*(2), 203-231.

Anderson, M. C., & Hanslmayr, S. (2014). Neural mechanisms of motivated forgetting. *Trends in cognitive sciences*, *18*(6), 279-292.

Asefi, S. L. & Garry, M. (2003). Absolut® memory distortions: alcohol placebos influence the misinformation effect. *Psychological Science*, 14(1), 77-80.

Atance, C. M. & O'Neill, D. K. (2001). Episodic future thinking. *Trends in Cognitive Sciences, 5*(12), 533-539.

Audi, R. (1995). Memorial justification. *Philosophical Topics*, *23*(1), 31-45.

Barrett, C. & Kurzban, R. (2006). Modularity in cognition: framing the debate. *Psychological Review*, *113*(3), 628-647.

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

Basu, S., Dickhaut, J., Hecht, G., Towry, K. & Waymire, G. (2009). Recordkeeping alters economic history by promoting reciprocity. *Proceedings of the National Academy of Sciences*, *106*(4), 1009-1014.

Bateson, M., Nettle, D. & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters*, *2*, 412-414.

Bauer, P. J. & Leventon, J. S. (2013). Memory for one-time experiences in the second year of life: implications for the status of episodic memory. *Infancy, 18*(5), 755-781.

Bell, B. E. & Loftus, E. F. (1988). Degree of detail of eyewitness testimony and mock juror judgments. *Journal of Applied Social Psychology*, *18*(14), 1171-1192.

Bell, B. E. & Loftus, E. F. (1989). Trivial persuasion in the courtroom: the power of (a few) minor details. *Attitudes and Social Cognition*, *56*(5), 669-679.

Bell, R., Koranyi, N., Buchner, A. & Rothermund, K. (2017). The implicit cognition of reciprocal exchange: automatic retrieval of positive and negative experiences with partners in a prisoner's dilemma game. *Cognition and Emotion*, *31*(4), 657-670.

Bell, R., Schain, C., & Echterhoff, G. (2014). How selfish is memory for cheaters? Evidence for moral and egoistic biases. *Cognition*, *132*(3), 437-442.

Benney, K. S. & Henkel, L. A. (2006). The role of free choice in memory for past decisions. *Memory*, *14*(8), 1001-1011.

Benoit, R. G., & Schacter, D. L. (2015). Specifying the core network supporting episodic simulation and episodic memory by activation likelihood estimation. *Neuropsychologia*, *75*, 450-457.

Berliner, D.C. (2005). The abuses of memory: Reflections on the memory boom in anthropology. *Anthropological Quarterly, 78*(1), 197-211.

Bernecker, S. (2008). *The metaphysics of memory*. Springer.

Bernecker, S. (2010). *Memory: A philosophical study*. Oxford University Press.

Berntsen, D. (2009). *Involuntary autobiographical memories: An introduction to the unbidden past*. Cambridge University Press.

Bietti, L. M., Tilston, O., & Bangerter, A. (2018). Storytelling as adaptive collective sensemaking. *Topics in Cognitive Science*. Doi: https://doi.org/10.1111/tops.12358

Blank, H., & Launay, C. (2014). How to protect eyewitness memory against the misinformation effect: A meta-analysis of post-warning studies. *Journal of Applied Research in Memory and Cognition*, *3*(2), 77-88.

Blank, H. (2009). Remembering: A theoretical interface between memory and social psychology. *Social Psychology*, *40*(3), 164-175.

Boehm, C. (2012). *Moral origins: The evolution of virtue, altruism and shame.* Basic Books.

Borst, G., & Kosslyn, S. M. (2008). Visual mental imagery and visual perception: Structural equivalence revealed by scanning processes. *Memory & Cognition*, *36*(4), 849-862.

Boyer, P. (2008). Evolutionary economics of mental time travel? *Trends in cognitive sciences*, *12*(6), 219-224.

Boyer, P. (2009). What are memories for? Functions of recall in cognition and culture. In: P. Boyer & J. V. Wertsch (Eds.), *Memory in mind and culture* (pp. 2-28). Cambridge University Press.

Brandom, R. (1983). Asserting. *Noûs*, 637-650.

Bregman, N. J. & McAllister, H. A. (1982). Eyewitness testimony: The role of commitment in increasing reliability. *Social Psychology Quarterly*, *45*(3), 181-184.

Breithaupt, F. (2011). The birth of narrative out of the spirit of the excuse: A speculation. *Poetics Today, 31*(1), 107-128.

Bright-Paul, A., Jarrold, C., & Wright, D. B. (2005). Age-appropriate cues facilitate source-monitoring and reduce suggestibility in 3-to 7-year-olds. *Cognitive Development*, *20*(1), 1-18.

Brosnan, S. F. & DeWaal, F. B. (2002). A proximate perspective on reciprocal altruism. *Human Nature*, *13*(1), 129-152.

Brown, A. S., & Murphy, D. R. (1989). Cryptomnesia: Delineating inadvertent plagiarism. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(3), 432-442.

Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in Cognitive Sciences*, *11*(2), 49-57.

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network. *Annals of the New York Academy of Sciences*, *1124*(1), 1-38.

Burge, T. (1993). Content preservation. *The Philosophical Review*, *102*(4), 457-488.

Burgess, P. W. & Shallice, T. (1996). Confabulation and the control of recollection. *Memory,4*(4), 359-411.

Burns, P., Russell. C. & Russell, J. (2015). Pre-school children's proto-episodic memory assessed by deferred imitation. *23*(8), 1172-1192.

Byrne, R. M. (2016). Counterfactual thought. *Annual Review of Psychology*, *67*, 135-157.

Campbell, J. (1996). Human vs. animal time. In: M.A. Pastor and J. Artieda (eds.), *Time, Internal Clocks and Movement* (pp. 115-126). Elsevier.

Carruthers, P. (2009). How we know our own minds: The relationship between mindreading and metacognition. *Behavioral and Brains Sciences,* 32.

Castelain, T., Bernard, S., Van der Henst, J. B., & Mercier, H. (2016). The influence of power and reason on young Maya children's endorsement of testimony. *Developmental science*, *19*(6), 957-966.

Chater, N. & Oaksford (2013). Programs as causal models: Speculations on mental programs and mental representation. *Cognitive Science, 37*, 1171-1191.

Cheng, S. & Werning, M. (2016). What is episodic memory if it is a natural kind? *Synthese, 193*(5), 1345-1385.

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

Clark-Foos, A., Brewer, G. & Marsh, R. L. (2015). Judging the reality of others' memories. *Memory*, 23(3), 427-436.

Clayton, N. S. & Dickinson, A. (1998). Episodic-like memory during cache recovery by scrub jays. *Nature*, *395*, 272-4.

Clayton, N. S. & Russell (2009). Looking for episodic memory in animals and young children: prospects for a new minimalism. *Neuropsychologia*, *47*, 2330-2340.

Clément, F., Koenig, M., & Harris, P. (2004). The ontogenesis of trust. *Mind & Language*, *19*(4), 360-379.

Clifasefi, S. L., Garry, M., Harper, D. N., Sharman, S. J. & Sutherland, R. (2007). Psychotropic placebos create resistance to the misinformation effect. *Psychonomic Bulletin & Review*, *14*(1), 112-117.

Cochran, K. J., Greenspan, R. L., Bogart, D. F., & Loftus, E. F. (2016). Memory blindness: Altered memory reports lead to distortion in eyewitness memory. *Memory & Cognition*, *44*(5), 717-726.

Coman, A., Momennejad, I., Drach, R. D., & Geana, A. (2016). Mnemonic convergence in social networks: The emergent properties of cognition at a collective level. *Proceedings of the National Academy of Sciences*, *113*(29), 8171-8176.

Coman, A., Stone, C. B., Castano, E., & Hirst, W. (2014). Justifying atrocities: The effect of moral-disengagement strategies on socially shared retrieval-induced forgetting. *Psychological Science*, *25*(6), 1281-1285.

Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological review*, *107*(2), 261-288.

Conway, M. A. (2005). Memory and the self. *Journal of Memory and Language*, *53*, 594-628.

Cosmides, L. & Tooby, J. (2000). Consider the source: The evolution of adaptations for decoupling and metarepresentation. In: D. Sperber (ed.), *Metarepresentations: A Multidisciplinary Perspective*. (pp. 53-115). Oxford University Press.

Craver, C. F., Kwan, D., Steindam, C. & Rosenbaum, R. S. (2014a). Individuals with episodic amnesia are not stuck in time. *Neuropsychologia*, *57*, 191-195.

Craver, C. F., Cova, F., Green, L., Myerson, J., Rosenbaum, R. S. et al. (2014b). An allais paradox without mental time travel. *Hippocampus*, *24*, 1375-1380.

Csibra, G. & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society: B, 366*(1567), 1149-57.

Dawkins, R. & Krebs, J. R. (1978). Animal signals: information or manipulation. In J. R. Krebs and N.B. Davies (Eds.), *Behavioral ecology: an evolutionary approach* (pp. 282-309). Basil Blackwell Scientific Publications.

Dab, S., Claes, T., Morais, J., & Shallics, T. (1999). Confabulation with a selective descriptor impairment. *Cognitive Neurpsychology, 16*(3-5), 215-242.

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

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

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

Debus, D. (2014). 'Mental time travel': Remembering the past, imagining the future, and the particularity of events. *Review of Philosophy and Psychology*, *5*(3), 333-350.

Demiray, B., Mehl, M. R., & Martin, M. (2018). Conversational Time Travel: Evidence of a Retrospective Bias in Real Life Conversations. *Frontiers in Psychology*, *9*. doi: 10.3389/fpsyg.2018.02160

Dennett, D.C. (2009). The part of cognitive science that is philosophy. *Topics in Cognitive Science, 1*, 231-236.

Desalles, J-L. (2007a). *Why we talk: the evolutionary origins of language*. Oxford University Press.

Desalles, J-L. (2007b). Storing events to retell them. *Behavioral and Brain Sciences*, 30(3), 321-322.

DeScioli, P., & Kurzban, R. (2009). Mysteries of morality. *Cognition*, *112*(2), 281-299.

Dew, I. T. Z. & Cabeza, R. (2011). The porous boundaries between explicit and implicit memory: behavioral and neural evidence. *Annals of the New York Academy of Sciences, 1224,* 174-190.

Dodd, D. H., & Bradshaw, J. M. (1980). Leading questions and memory: pragmatic constraints. *Journal of Verbal Learning and Verbal Behavior*, 19, 695-704.

Dokic, J. (2001). Is memory purely preservative? In: C. Hoerl & T. McCormack (Eds.), *Time and Memory: Issues in philosophy and psychology* (pp. 213-232). Oxford University Press.

Dokic, J. (2014). Feeling the past: A two-tiered account of episodic memory. *Review of Philosophy and Psychology*, *5*(3), 413-426.

Drummey, A. B., & Newcombe, N. S. (2002). Developmental changes in source memory. *Developmental Science*, *5*(4), 502-513.

Dunbar, R. I. M. (2004). Gossip in evolutionary perspective. *Review of General Psychology*, *8*(2), 100-110.

Dunbar, R. (1998). *Grooming, gossip, and the evolution of language*. Harvard University Press.

Dunbar, R. I. M., Duncan, N. D. C., & Nettle, D. (1997). Human conversational behavior. *Human Nature*, *6*, 67-78.

Eagly, A. H., Kulesa, P., Chen, S. & Chaiken, S. (2001). Do attitudes affect memory? Tests of the congeniality hypothesis. *Current Directions in Psychological Science*, *10*(1), 5-9.

Eggins, S., & Slade, D. (2005). *Analysing casual conversation*. Equinox Publishing Ltd.

Eich, E., Nelson, A. L., Leghari, M. A., & Handy, T. C. (2009). Neural systems mediating field and observer memories. *Neuropsychologia*, *47*(11), 2239-2251.

Eichenbaum, H., & Cohen, N. J. (2014). Can we reconcile the declarative memory and spatial navigation views on hippocampal function? *Neuron*, *83*(4), 764-770.

Elekes, F., Varga, M., & Király, I. (2016). Evidence for spontaneous level-2 perspective taking in adults. *Consciousness and Cognition*, *41*, 93-103.

Engelmann, J. M., Hermann, E. & Tomasello, M. (2012). Five-year olds, but not chimpanzees, attempt to manage their reputations. *PLOSone*, *7*(10), e48433.

Fernandez, J. (2018). The functional character of memory. In: D. Debus, K. Michaelian and D. Perrin (eds.). *New Directions in the Philosophy of Memory* (pp. 52-72). Routledge.

Fessler, D. M., & Quintelier, K. (2013). Suicide bombers, weddings, and prison tattoos: An evolutionary perspective on subjective commitment and objective commitment. In K. Sterelny, R. Joyce, B. Calcott, R B. Fraser (Eds.). *Cooperation and its evolution*. (pp. 459-484). MIT Press.

Firestone, C. & Scholl, B. J. (2016). Cognition does not affect perception: evaluating the evidence for so-called 'top-down' effects. *Behavioral and Brain Sciences*, 39.

Fitzgerald, J. M. & Broadbridge, C. L. (2013). Latent constructs of the autobiographical memory questionnaire: a recollection-belief model of autobiographical experience. *Memory*, *21*(2), 230-248.

Fivush, R. & Bauer, P. J. (2010). The emergence of recollection: how we learn to recall ourselves in the past. In J. H. Mace (Ed.). *The Act of Remembering: toward an understanding of how we recall the past*. (pp. 259-284). Wiley-Blwackwell.

Foster, E. K. (2004). Research on gossip: Taxonomy, methods, and future directions. *Review of General Psychology*, *8*(2), 78-99.

French, L., Garry, M. & Mori, K. (2011). Relative – not absolute – judgments of credibility affect susceptibility to misinformation conveyed during discussion. *Acta Psychologica*, 136, 119-128.

Freundlieb, M., Kovács, Á. M., & Sebanz, N. (2018). Reading your mind while you are reading: evidence for spontaneous visuospatial perspective taking during a semantic categorization task. *Psychological Science*, *29*(4), 614-622.

Gabbert, F., Memon, A. & Wright, D. B. (2007). I saw it for longer than you: the relationship between perceived encoding duration and memory conformity. *Acta Psychologica*, 124, 319-331.

Gaesser, B., Spreng, R. N., McLelland, V. C., Addis, D. R., & Schacter, D. L. (2013). Imagining the future: evidence for a hippocampal contribution to constructive processing. *Hippocampus*, *23*(12), 1150-1161.

Gardiner, J. M. (2001). Episodic memory and autonoetic consciousness: a first–person approach. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *356*(1413), 1351-1361.

Garry, M., Manning, C. G., Loftus, E. F., & Sherman, S. J. (1996). Imagination inflation: Imagining a childhood event inflates confidence that it occurred. *Psychonomic Bulletin & Review*, *3*(2), 208-214.

Gervais, M. M., & Fessler, D. M. (2017). On the deep structure of social affect: Attitudes, emotions, sentiments, and the case of "contempt". *Behavioral and Brain Sciences*, *40*.

Gluckman, M. (1973). *The judicial process among the Barotse of Northern Rhodesia*. Manchester University Press.

Giardini, F. & Conte, R. (2011). Gossip for social control in natural and artificial societies. *Simulation*, *0*(00), 1-15.

Gilbert, D.T. & Malone, P.S. (1995). The correspondence bias. *Psychological Bulletin, 117*, 21-38.

Giles, J. W., Gopnik, A., & Heyman, G. D. (2002). Source monitoring reduces the suggestibility of preschool children. *Psychological Science*, *13*(3), 288-291.

Goethals, G. R. & Reckman. R. F. (1973). The perception of consistency in attitudes. *Journal of Experimental Social Psychology*, *9*, 491-501.

Goldman, A. I. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. Oxford University Press.

Goodman, N. (1983). *Fact, fiction and forecast*. Harvard University Press.

Gopnik, A., Glymour, C. Sobel, D.M., Schulz, L. et al. (2004). A theory of causal learning in children: Causal maps and bayes nets. *Psychological Review, 111*(1), 3-32.

Gopnik, A., & Graf, P. (1988). Knowing how you know: Young children's ability to identify and remember the sources of their beliefs. *Child Development*, 1366-1371.

Gopnik, A. (2000). Explanation as orgasm: The function, evolution and phenomenology of the theory-formation system. In F. Keil & R. Wilson (eds.). *Cognition and Explanation*. (pp. 299-323) MIT Press.

Greene, E. (1981). Whodunit? Memory for evidence in text. *American Journal of Psychology, 94*(3), 479-496.

Gupta, A. S., van der Meer, M. A. A., Touretzky, D. S. & Redish, A. D. (2010). Hippocampal replay is not a simple function of experience. *Neuron*, *65*, 695-705.

Haigh, S. N., & Robinson, E. J. (2009). What children know about the source of their knowledge without reporting it as the source. *European Journal of Developmental Psychology*, *6*(3), 318-336.

Hannula, D. E., Baym, C. L., Warren, D. E., & Cohen, N. J. (2012). The eyes know: eye movements as a veridical index of memory. *Psychological Sciences, 23*(3), 278-287.

Hannula, D. E. & Greene, A. J. (2012). The hippocampus reevaluated in unconscious learning and memory: at a tipping point? *Frontiers in Human Neuroscience*, *6*, 80.

Hannula, D. E. & Ranganath, C. (2009). The eyes have it: Hippocampal activity predicts expression of memory in eye movements. *Neuron, 63*, 592-599.

Hassabis, D., & Maguire, E. A. (2007). Deconstructing episodic memory with construction. *Trends in Cognitive Sciences*, *11*(7), 299-306.

Hassabis, D., Kumaran, D., Vann, S. D., & Maguire, E. A. (2007). Patients with hippocampal amnesia cannot imagine new experiences. *Proceedings of the National Academy of Sciences*, *104*(5), 1726-1731.

Hassabis, D., & Maguire, E. A. (2009). The construction system of the brain. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1521), 1263-1271.

Haviland, J. B. (1977). *Gossip, reputation and knowledge in Zinacantan*. University of Chicago Press.

Hayne, H., & Jack, F. (2011). Childhood amnesia. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(2), 136-145.

Henke, K. (2010). A model for memory systems based on processing modes rather than consciousness. *Nature Reviews Neuroscience, 11*, 523-532.

Helming, K.A., Strickland, B. & Jacob, P. (2014). Making sense of early false-belief understanding. *Trends in Cognitive Sciences*, *18*(4), 167-170.

Hemmer, P., & Steyvers, M. (2009). A Bayesian account of reconstructive memory. *Topics in Cognitive Science*, *1*(1), 189-202.

Henkel, L. A. & Mather, M. (2007). Memory attributions for choices: how beliefs shape our memories. *Journal of Memory and Language*, *57*(2), 163-176.

Hess, N.H. & Hagen, E.H. (2006). Psychological adaptations for assessing gossip veracity. *Human Nature, 17*(3), 337-354.

Higgins, E. T. & Rholes, W. S. (1978). 'Saying is believing': effects of message modification on memory and liking for the person described. *Journal of Experimental Social Psychology*, *14*, 363-378.

Hills, T. T. & Butterfill, S. (2015). From foraging to autonoetic consciousness: The primal self as a consequence of embodied prospective foraging. *Current Zoology, 61*(2), 368-381.

Hirst, W. & Echterhoff, G. (2012). Remembering in conversation: The social sharing and reshaping of memories. *Annual Review of Psychology, 63*(21), 21.1-21.25.

Hirst, W., Yamashiro, J. K. & Coman, A. (2018). Collective memory from a psychological perspective. *Trends in Cognitive Sciences, 22*(5), 438-451.

Hoerl, C. & McCormack, T. (forthcoming). Thinking in and about time: a dual systems perspective on temporal cognition. *Behavioral and Brain Sciences.*

Hoerl, C. & McCormack, T. (2016). Making decisions about the future: Regret and the cognitive function of episodic memory. In K. Michaelian, S. B. Klein, & K. K. Szpunar (Eds.), *Seeing the future: Theoretical perspectives on future-oriented mental time travel* (pp. 241-266). Oxford University Press.

Irish, M., Addis, D. R., Hodges, J. R., & Piguet, O. (2012). Considering the role of semantic memory in episodic future thinking: Evidence from semantic dementia. *Brain*, *135*(7), 2178-2191.

Jablonka, E. (2017). Collective narratives, false memories, and the origins of autobiographical memory. *Biology & Philosophy*, *32*(6), 839-853.

Jaeger, A., Lauris, P., Selmeczy, D. & Dobbins, I. G. (2012). The costs and benefits of memory conformity. *Memory & Cognition,* 40, 101-112.

Jing, H. G., Madore, K. P., & Schacter, D. L. (2017). Preparing for what might happen: an episodic specificity induction impacts the generation of alternative future events. *Cognition*, *169*, 118-128.

Jing, H. G., Madore, K. P., & Schacter, D. L. (2016). Worrying about the future: An episodic specificity induction impacts problem solving, reappraisal, and well-being. *Journal of Experimental Psychology: General*, *145*(4), 402.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*(1), 3.

Johnson, M. K., Foley, M. A., Suengas, A. G., & Raye, C. L. (1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events. *Journal of Experimental Psychology: General*, *117*(4), 371.

Johnson, M. K., Nush, J. G. & Mitchell, K. J. (1998) Interpersonal reality monitoring: judging the sources of other people's memories. *Social Cognition,* 16(2), 199-224.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, *88*(1), 67.

Johnson, M. K. & Suengas, A. G. (1989). Reality monitoring judgments of other people's memories. *Bulletin of the Psychonomic Society*, 27(2), 107(110).

Johnson, M. K. (1991). Reality monitoring: Evidence from confabulation in organic brain disease patients. In G. P. Prigatano & D. L. Schacter (Eds.), *Awareness of deficit after brain injury: Clinical and theoretical issues* (pp. 176-197). New York, NY, US: Oxford University Press.

Johnson, M. K. (2005). The relation between source memory and episodic memory: comment on Siedlicki et al. (2005). *Psychology and Aging*, 20(3), 529-531.

Kappes, A., & Crockett, M. J. (2016). The benefits and costs of a rose-colored hindsight. *Trends in Cognitive Sciences*, *20*(9), 644-646.

Keven, N., Kurczek, J., Rosenbaum, R. S., & Craver, C. F. (2018). Narrative construction is intact in episodic amnesia. *Neuropsychologia*, *110*, 104-112.

Keven, N. (2016). Events, narratives and memory. *Synthese*, *193*(8), 2497-2517.

Keven, N. (2018). Carving event and episodic memory at their joints. *Behavioral and Brain Sciences*, *41*.

Kim, S., Dede, A. J., Hopkins, R. O., & Squire, L. R. (2015). Memory, scene construction, and the human hippocampus. *Proceedings of the National Academy of Sciences*, *112*(15), 4767-4772.

Király, I., Oláh, K., Csibra, G., & Kovács, A. (2018). Retrospective attribution of false beliefs in 3-year-old children. *Proceedings of the National Academy of Sciences, 115*(45), 11477-11482.

Kleimann-Weiner, M., Ho, M. K., Austerweil, J. L., Littman,M. L. & Tennenbaum, J. B. (2016). Coordinate to cooperate or compete: Abstract goals and joint intentions in social interaction. *COSGSCI 2016*. Available online at: https://par.nsf.gov/biblio/10026426

Klein, S. B., Cosmides, L., Gangi, C. E., Jackson, B., Tooby, J., & Costabile, K. A. (2009). Evolution and episodic memory: An analysis and demonstration of a social function of episodic recollection. *Social Cognition*, *27*(2), 283-319.

Klein, S. B., Cosmides, L., Tooby, J., & Chance, S. (2002). Decisions and the evolution of memory: multiple systems, multiple functions. *Psychological review*, *109*(2), 306.

Klein, S. B., German, T. P., Cosmides, L., & Gabriel, R. (2004). A theory of autobiographical memory: Necessary components and disorders resulting from their loss. *Social Cognition*, *22*(5), 460-490.

Klein, S. B., Loftus, E. & Kihlstrom, J. F. (2002) Memory and temporal experience: The effects of episodic memory loss on an amnesic patient's ability to remember the past and imagine the future. *Social Cognition, 20*(5)*,* 353–79.

Klein, S. B. & Nichols, S. (2012). Memory and the sense of personal identity. *Mind*, 121(483), 677-702.

Klein, S. B (2013a). Making the case that episodic recollection is attributable to operations occurring at retrieval rather than to content stored in a dedicated subsystem of long-term memory. *Frontiers in behavioral neuroscience*, *7*, 1-14.

Klein, S. B. (2013b). The complex act of projecting oneself into the future. *WIREs Cognitive Science*, 4, 63-79.

Klein, S. B. (2013c). Autonoesis and belief in a personal past: an evolutionary theory of memory indices. *Review of Philosophy and Psychology*, *5*, 417-447.

Klein, S. B. (2014). Autonoesis and belief in a personal past: An evolutionary theory of episodic memory indices. *Review of Philosophy and Psychology, 5*, 427-447.

Klein, S. (2015). What memory is. *Wiley Interdisciplinary Reviews: Cognitive Science, 6*(1), 1-38.

Koriat, A., & Goldsmith, M. (1996). Memory metaphors and the real-life/laboratory controversy: Correspondence versus storehouse conceptions of memory. *Behavioral and Brain Sciences*, *19*(2), 167-188.

Kornblith, H. (2012). *On reflection*. Oxford University Press.

Kouchaki, M., & Gino, F. (2016). Memories of unethical actions become obfuscated over time. *Proceedings of the National Academy of Sciences*, *113*(22), 6166-6171.

Kraut, R. E. (1978). Verbal and nonverbal cues in the perception of lying. *Journal of Personality and Social Psychology*, *36*(4), 380-391.

Krebs, J. R. & Dawkins, R. (1984). Animal signals: mind-reading and manipulation. In J. R. Krebs and N. B. Davies (eds.), *Behavioral ecology: an evolutionary approach*, 2nd edition, 380-402.

Kulakova, E., Aichhorn, M., Schurz, M., Kronbichler, M., & Perner, J. (2013). Processing counterfactual and hypothetical conditionals: An fMRI investigation. *NeuroImage*, *72*, 265-271.

Kwan, D., Craver, C. F., Green, L., Myerson, J., Boyer, P. & Rosenbaum, S. (2012). Future decision-making without episodic mental time travel. *Hippocampus*, *22*(6), 1215-1219.

Kwong See, S. T., Wood. T. L. & Hoffman, H. G. (2001). Perceptions of an old female eyewitness: is the older eyewitness believable? *Psychology and Aging*, 16(2), 346-350.

Laeng, B. & Sulutvedt, U. (2014). The eye pupil adjusts to imaginary light. *Psychological Science, 25*(1), 188-197.

Lambek, M. (1996). The past imperfect: remembering as moral practice. In: P. Antze & M. Lambek (eds.), *Tense past: Cultural Essays in Trauma and Memory* (pp. 235-255). Routledge.

Lampinen, J. M. & Smith, V. L. (1995). The incredible (and sometimes incredulous) child witness: Child eyewitnesses' sensitivity to source credibility cues. *Journal of Applied Psychology*, *80*(5), 621-627.

Lampinen, J. M., Neuschatz, J. S. & Payne, D. G. (1997). Memory illusions and consciousness: examining the phenomenology of true and false memories. *Current Psychology*, *16*(3), 181-224.

Leding, J. K. (2012). False memories and persuasion strategies. *Review of General Psychology*, *16*(3), 256-268.

Leimgruber, K. L., Shaw, A., Santos, L. R. & Olson, K. R. (2012). Young children are more generous when others are aware of their actions. *PLOSone*, *7*(10), e48292.

Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind.". *Psychological Review*, *94*(4), 412.

Levine, L. J. (1997). Reconstructing memory for emotions. *Journal of Experimental Psychology: General, 126*(2), 165-177.

Lindsay, D. S. & Johnson, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory & Cognition*, 17(3), 349-358.

Loftus, E. F., & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals*, *25*(12), 720-725.

Loftus, E. (1977). Shifting human color memory. *Memory & Cognition*, *5*(6), 696-699.

Loftus, E. (2005). Planting misinformation in the human mind: a 30-year investigation of the malleability of memory. *Learning & Memory, 12*(4), 361-366.

Ludmer, R., Edelsson, M. G. & Dudai, Y. (2015). The naïve and the distrustful: state dependency of hippocampal computations in manipulative memory distortion. *Hippocampus,* 25, 240-252.

MacDonald, C. J., Lepage, K. Q., Eden, U. T., & Eichenbaum, H. (2011). Hippocampal "time cells" bridge the gap in memory for discontiguous events. *Neuron*, *71*(4), 737-749.

Madore, K. P., Addis, D. R., & Schacter, D. L. (2015). Creativity and memory: Effects of an episodic-specificity induction on divergent thinking. *Psychological Science*, *26*(9), 1461-1468.

Madore, K. P., Gaesser, B., & Schacter, D. L. (2014). Constructive episodic simulation: Dissociable effects of a specificity induction on remembering, imagining, and describing in young and older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(3), 609.

Madore, K. P., Jing, H. G., & Schacter, D. L. (2018). Selective effects of specificity inductions on episodic details: evidence for an event construction account. *Memory*, 1-11.

Madore, K. P., & Schacter, D. L. (2014). An episodic specificity induction enhances means-end problem solving in young and older adults. *Psychology and aging*, *29*(4), 913.

Maguire, E. A., & Mullally, S. L. (2013). The hippocampus: A manifesto for change. *Journal of Experimental Psychology: General*, *142*(4), 1180-1189.

Maguire, E. A., Intraub, H., & Mullally, S. L. (2016). Scenes, spaces, and memory traces: What does the hippocampus do?. *The Neuroscientist*, *22*(5), 432-439.

Maguire, E. A., Vargha-Khadem, F., & Hassabis, D. (2010). Imagining fictitious experiences: evidence from developmental amnesia. *Neuropsychologia*, *48*, 3187-3192.

Mahr, J.B., Mercier, H., Mascaro, O., & Csibra, G. (2019). The effect of disagreement on children's source memory performance. *Mansuscript in preparation.*

Marr, D. (1982). Vision: a computational investigation into the human representation and processing of visual information. *MIT Press.*

Martin, C.B. & Deutscher, M. (1966). Remembering. *The Philosophical Review, 75*(2), 161-196.

Martin-Ordas, G., Haun, D., Colmenares, F. & Call, J. (2010). Keeping track of time: evidence for episodic-like memory in great apes. *Animal Cognition*, *13*, 331-340.

Mascaro, O., & Sperber, D. (2009). The moral, epistemic, and mindreading components of children's vigilance towards deception. *Cognition*, *112*(3), 367-380.

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

Mather, M., Shafir, E. & Johnson, M. K. (2000). Misremembrance of options past: source monitoring and choice. *Psychological Science*, *11*(2), 132-138.

Mather, M., Shafir, E. & Johnson, M. K. (2003). Remembering chosen and assigned options. *Memory & Cognition*, *31*(3), 422-433.

Mather, M. & Johnson, M. K. (2000). Choice-supportive source monitoring: do our decisions seem better to us as we age? *Psychology and Aging*, *15*(4), 596-606.

Mazzarella, D., Reinecke, R., Noveck, I., & Mercier, H. (2018). Saying, presupposing and implicating: How pragmatics modulates commitment. *Journal of Pragmatics, 133*, 15-27.

Mazzoni, G. & Kirsch, I., (2002). Autobiographical memories and beliefs: a preliminary metacognitive model. In T.J. Perfect & B.L. Schwartz (Eds.), *Applied metacognition*. (pp. 121-145) Cambridge University Press.

McClelland, J. L., McNaughton, B. L., & O'reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419-457.

McDonough, I. M., & Gallo, D. A. (2010). Separating past and future autobiographical events in memory: Evidence for a reality monitoring asymmetry. *Memory & Cognition*, *38*(1), 3-12.

McDonough, I. M., & Gallo, D. A. (2013). Impaired retrieval monitoring for past and future autobiographical events in older adults. *Psychology and aging*, *28*(2), 457.

McMyler, B. (2007). Knowledge at second hand. *Inquiry*, *50*(5), 511-540.

McTaggart, J. E. (1908). The unreality of time. *Mind*, 457-474. Michaelian, K. & Robins, S. (2018). Beyond the causal theory? Fifty years after Martin and Deutscher. In: K. Michaelian, D. Debus and D. Perrin (eds.). *New Directions in the Philosophy of Memory* (pp. 21-40). Routledge.

Mercier, H. & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*(2), 57-74.

Mercier, H. & Sperber, D. (2017). *The enigma of reason*. Harvard University Press.

Mercier, H. (2016). The argumentative theory: predictions and empirical evidence. *Trends in Cognitive Sciences*, *20*(9), 689-700.

Mercier, H. (2017). How gullible are we? A review of the evidence from psychology and social science. *Review of General Psychology, 21*(2), 103-123.

Merckelbach, H., Jelcic, M. & Pieters, M. (2010). The residual effect of feigning: how intentional faking may evolve into a less conscious form of symptom reporting. *Journal of Experimental and Clinical Neuropsychology*, *33(*1), 131-139.

Merckelbach, H., Jelcic, M. & Pieters, M. (2011). Misinformation increases symptom reporting: a test-retest study. *Journal of the Royal Society of Medicine Short Reports*, *2*(75), 1-6.

Michaelian, K. & Robins, S. (2018). Beyond the causal theory? Fifty years after Martin and Deutscher. In: K. Michaelian, D. Debus and D. Perrin (eds.). *New Directions in the Philosophy of Memory* (pp. 21-40). Routledge.

Michaelian, K. (2016). *Mental time travel*: *Episodic memory and our knowledge of the personal past.* MIT Press.

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

Michaelian, K. (2012a). Metacognition and endorsement. *Mind & Language*, 27(3), 284-307.

Michaelian, K. (2012b). (Social) metacognition and (self-)trust. *Review of Philosophy and Psychology*, 3, 481-514.

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

Michaelian, K. (2011b). Generative memory. *Philosophical Psychology, 24*(3), 323-342.

Milner, B. (1962). Les troubles de la mémoire accompagnant des lésions hippocampiques bilatérales. In: *Physiologie de l'hippocampe* (pp. 257–272). Centre National de la Recherche Scientifique. English translation: B. Milner & S. Glickman (Eds., 1965). (pp. 97–111). Van Nostrand.

Mitchell, K. J., & Johnson, M. K. (2009). Source monitoring 15 years later: What have we learned from fMRI about the neural mechanisms of source memory? *Psychological Bulletin*, *135*(4), 638.

Millikan, R. G. (1984). *Language, thought and other biological categories: new foundations for realism*. MIT Press.

Moscovitch, M., & Nadel, L. (1998). Consolidation and the hippocampal complex revisited: In defense of the multiple-trace model. *Current Opinion in Neurobiology*, *8*(2), 297-300.

Moscovitch, M. (2008). The hippocampus as a "stupid", domain-specific module: implications for theories of recent and remote memory, and of imagination. *Canadian Journal of Experimental Psychology*, 62(1), 62-79.

Moulton, S. T., & Kosslyn, S. M. (2009). Imagining predictions: Mental imagery as mental emulation. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1521), 1273-1280.

Mullally, S. & Maguire, E. A. (2014). Learning to remember: The early ontogeny of episodic memory. *Developmental Cognitive Neuroscience, 9*, 12-29.

Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology*, *7*(2), 217-227.

Nagel, J. (2015). The social value of reasoning in epistemic justification. *Episteme*, *12*(2), 297-308.

Nanay, B. (2015). Perceptual content and the content of mental imagery. *Philosophical Studies*, *172*(7), 1723-1736.

Nagy, D. G. & Orban, G. (2016). Episodic memory as a prerequisite for online updating of model structure. *Proceedings of the 38th Annual Conference of the Cognitive Science Society.*

Nash, R. A., Wheeler, R. L. & Hope, L. (2014). On the persuability of memory: is changing people's memories no more than changing their minds? *British Journal of Psychology*, *106*(2), 308-326.

Nelson, K. (1993). The psychological and social origins of autobiographical memory. *Psychological Science*, *4*(1), 7-14.

Nelson, K. & Fivush, R. (2004). The emergence of autobiographical memory: a social cultural developmental theory. *Psychological Review*, *111*(2), 486-511.

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

Nigro, G., & Neisser, U. (1983). Point of view in personal memories. *Cognitive psychology*, *15*(4), 467-482.

Norman, D. A., & Bobrow, D. G. (1979). Descriptions: An intermediate stage in memory retrieval. *Cognitive Psychology*, *11*(1), 107-123.

Nowak, M.A. & Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. *Nature, 393,* 573-577.

Nowak, M.A. & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature, 437*, 1291-1298.

Nyberg, L., Kim, A. S., Habib, R., Levine, B., & Tulving, E. (2010). Consciousness of subjective time in the brain. *Proceedings of the National Academy of Sciences*, 201016823.

Olick, J. K., & Levy, D. (1997). Collective memory and cultural constraint: Holocaust myth and rationality in German politics. *American Sociological Review*, 62(6), 921-936.

Olick, J. K. (1999). Collective memory: The two cultures. *Sociological theory*, *17*(3), 333-348.

Olsen, R. K., Moses, S. N., Riggs, L. & Ryan, J. D. (2012). The hippocampus supports multiple cognitive processes through relational binding and comparison. *Frontiers in Human Neuroscience*, *6*(146).

O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Clarendon Press.

Parikh, N., Ruzic, L., Stewart, G. W., Spreng, R. N., & De Brigard, F. (2018). What if? Neural activity underlying semantic and episodic counterfactual thinking. *NeuroImage*, *178*, 332-345.

Pärnamets, P., Hall, L., & Johansson, P. (2015). Memory distortions resulting from a choice blindness task. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C.D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, pp. 1823-1828. Cognitive Science Society.

Pasupathi, M., McLean, K.C. & Weeks, T. (2009). To tell or not to tell: Disclosure and the narrative self. *Journal of Personality, 77*(1), 89-124.

Pasupathi, M., Stallworth, L. M. & Murdoch, K. (1998). How what we tell becomes what we know: listener effects on speakers' long-term memory for events. *Discourse Processes*, *26*(1), 1-25.

Payne, D. G., Elie, C. J., Blackwell, J. M., & Neuschatz, J. S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory and Language*, *35*(2), 261-285.

Pearson, J., Clifford, C. W., & Tong, F. (2008). The functional impact of mental imagery on conscious perception. *Current Biology*, *18*(13), 982-986.

Pearson, J., Naselaris, T., Holmes, E. A., & Kosslyn, S. M. (2015). Mental imagery: functional mechanisms and clinical applications. *Trends in Cognitive Sciences, 19*(10), 590-602.

Pearson, J. & Kosslyn, S. M. (2015). The heterogeneity of mental representation: ending the imagery debate. *Proceedings of the National Academy of Sciences*, *112*(33), 10089-92.

Perner, J. & Ruffman, T. (1995). Episodic memory and autonoetic consciousness: developmental evidence and a theory of childhood amnesia. *Journal of Experimental Child Psychology*, *59*, 516-548.

Perner, J., Kloo, D. & Stöttinger, E. (2007). Introspection & remembering. *Synthese*, *159*, 253-270.

Perner, J. (1991). *Understanding the representational mind*. MIT Press.

Perner, J. (2001). Episodic memory: essential distinctions and developmental implications. In: C. Moore & K. Lemmon (Eds.), *The self in time: developmental perspectives*, pp. 181-202. Lawrence Erlbaum Associates Publishers.

Perner, J. (2012). MiniMeta: in search of minimal criteria for metacognition. In: M. J. Beran, J. L. Brandl, J. Perner, & J. Proust (Eds.). *Foundations of metacognition*. (pp. 94-116). Oxford University Press.

Phillips, J., & Cushman, F. (2017). Morality constrains the default representation of what is possible. *Proceedings of the National Academy of Sciences*, *114*(18), 4649-4654.

Philips, I. (2014). Lack of imagination: individual differences in mental imagery and the significance of consciousness. In: M. Sprevak and J. Kallestrup (Eds.). *New Waves in Philosophy of Mind.* (pp. 278-300). Springer.

Piazza, J., Bering, J. M. & Ingram, G. (2011). “Princess Alice is watching you”: children’s belief in an invisible person inhibits cheating. *Journal of Experimental Child Psychology*, *109*, 311-320.

Pietraszewski, D. (2016). How the mind sees coalitional and group conflict: the evolutionary invariances of n-person conflict dynamics. *Evolution and Human Behavior*, *37*(6), 470-480.

Pillemer, D. B., Steiner, K. L., Kuwabara, K. J., Thomsen, D. K., & Svob, C. (2015). Vicarious memories. *Consciousness and Cognition*, *36*, 233-245.

Poole, R. (2008). Memory, history and the claims of the past. *Memory Studies*, *1*(2), 149-166.

Poppenk, J., Evensmoen, H. R., Moscovitch, M., & Nadel, L. (2013). Long-axis specialization of the human hippocampus. *Trends in Cognitive Sciences*, *17*(5), 230-240.

Prasada, S. (2000). Acquiring generic knowledge. *Trends in Cognitive Sciences, 4*, 66-72.

Proust, J. (2007). Metacognition and metarepresentation: is a self-directed theory of mind a precondition for metacognition? *Synthese*, *159*(2), 271-295.

Pylyshyn, Z. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences, 25*, 157-238.

Rimé, B., Mesquita, B., Boca, S. & Philippot, P. (1991). Beyond the emotional event: six studies on the social sharing of emotion. *Cognition & Emotion*, *5*(5-6), 435-465.

Rimé, B., Finkenauer, C., Luminet, O., Zech, E. & Phillippot (1998). Social sharing of emotions: new evidence and new questions. *European Review of Social Psychology*, *9*(1), 145-189.

Redshaw, J. (2014). Does metarepresentation make human mental time travel unique? *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(5), 519-531.

Roberts, R. P., Schacter, D. L., & Addis, D. R. (2017). Scene construction and relational processing: Separable constructs? *Cerebral Cortex*, *28*(5), 1729-1732.

Robins, S. (2016). Representing the past: Memory traces and the causal theory of memory. *Philosophical Studies, 173*(11), 2993-3013.

Robin, J., & Moscovitch, M. (2017). Details, gist and schema: hippocampal–neocortical interactions underlying recent and remote episodic and spatial memory. *Current Opinion in Behavioral Sciences*, *17*, 114-123.

Robinson, J. A., & Swanson, K. L. (1993). Field and observer modes of remembering. *Memory*, *1*(3), 169-184.

Robinson, E. J., & Whitcombe, E. L. (2003). Children's suggestibility in relation to their understanding about sources of knowledge. *Child Development*, *74*(1), 48-62.

Rodriguez, D. N. & Strange, D. (2015). False memories for dissonance inducing events. *Memory*, *23*(2), 203-212.

Roediger, H. L. (1996). Memory illusions. *Journal of Memory and Language*, *35*, 76-100.

Roediger, H. L. (2001). Why retrieval is the key process in understanding human memory. In: E. Tulving (Ed.), *Memory, consciousness and the brain: the Tallinn conference*, pp. 52-75.

Ross, L. (1977). "The intuitive psychologist and his shortcomings: Distortions in the attribution process". In Berkowitz, L. Advances in experimental social psychology 10 (pp. 173–220). Academic Press.

Ross, M., McFarland, C. & Fletcher, O. J. (1981). The effect of attitude on the recall of personal histories. *Journal of Personality and Social Psychology*, *40*(4), 627-634.

Ross, M, McFarland, C., Conway, M & Zanna, M. P. (1983). Reciprocal relation between attitudes and behavior recall. *Journal of Personality and Social Psychology*, *45*(2), 257-267.

Ross, M. (1989). Relation of implicit theories to the construction of personal histories. *Psychological Review*, *96*(2), 341-357.

Rubin, D. C., Schrauf, R. W. & Greenberg, D. L. (2003). Belief and recollection of autobiographical memories. *Memory & Cognition*, *32*(6), 887-901.

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

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

Sanitioso, R., Kunda, Z., & Fong, G. T. (1990). Motivated recruitment of autobiographical memories. *Journal of Personality and Social psychology*, *59*(2), 229.

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(5), 1255.

Schacter, D. L. (2001). *The Seven Sins of Memory: how the mind forgets and remembers*. Houghton Mifflin.

Schacter, D. L., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory: remembering the past and imagining the future. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *362*(1481), 773-786.

Schacter, D. L., & Addis, D. R. (2009). On the nature of medial temporal lobe contributions to the constructive simulation of future events. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1521), 1245-1253.

Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., & Szpunar, K. K. (2012). The future of memory: remembering, imagining, and the brain. *Neuron*, *76*(4), 677-694.

Schacter, D. L., Benoit, R. G., De Brigard, F., & Szpunar, K. K. (2015). Episodic future thinking and episodic counterfactual thinking: Intersections between memory and decisions. *Neurobiology of Learning and Memory*, *117*, 14-21.

Schacter, D. L., & Madore, K. P. (2016). Remembering the past and imagining the future: Identifying and enhancing the contribution of episodic memory. *Memory Studies*, *9*(3), 245-255.

Schacter, D.L. & Tulving, E. (1994). What are the memory systems of 1994? In: D.L. Schacter and E. Tulving (eds.). *Memory Systems* (pp. 1-38). MIT Press.

Schino, G. & Aureli, F. (2009). Reciprocal altruism in primates: Partner choice, cognition and emotions. *Advances in the Study of Behavior*, *39*, 45-69.

Schooler, J. W., Foster, R. A. & Loftus, E. F. (1988). Some deleterious consequences of the act of recollection. *Memory & Cognition*, *16*(3), 243-251.

Scoboria, A., Jackson, D. L., Talarico, J., Hanczakowski, M. Wzsman, L. & Mazzoni, G. (2014). The role of belief in occurrence within autobiographical memory. *Journal of Experimental Psychology: General*, *143*(3), 1242-1258.

Seemann, A. (2016). Reminiscing together: Joint experiences, epistemic groups, and sense of self. *Synthese*, 1-16. doi: https://doi.org/10.1007/s11229-016-1156-3

Shafto, P. & Goodman, N.D. (2008). Teaching games: Statistical sampling assumptions for learning in pedagogical situations. In *Proceedings of the 30th annual conference of the Cognitive Science Society* (pp. 1632-1637). Cognitive Science Society.

Shafto, P., Goodman, N.D. & Griffiths, T.L. (2014). A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cognitive Psychology, 71*, 55-89.

Sheldon, S., & Levine, B. (2016). The role of the hippocampus in memory and mental construction. *Annals of the New York Academy of Sciences*, *1369*(1), 76-92.

Sheldon, S. & Moscovitch, M. (2010). Recollective performance advantages for implicit memory tasks. *Memory*, *18*(7), 681-697.

Shtulman, A., & Carey, S. (2007). Improbable or impossible? How children reason about the possibility of extraordinary events. *Child development*, *78*(3), 1015-1032.

Sommerfed, G. D., Krambeck, H., Semmann, D. & Milinksi, M. (2007). Gossip as an alternative for direct observation in games of indirect reciprocity. *Proceedings of the National Academy of Sciences*, *104*(44), 17435-17440.

Speas, P. (2008). On the syntax and semantics of evidentials. *Language and Linguistics Compass*, *2*(5), 940-965.

Sperber, D., Clemént, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G. & Wilson, D. (2010). Epistemic Vigilance. *Mind & Language, 25*(4), 359-393.

Sperber, D. & Hirschfeld, L. A. (2004). The cognitive foundations of cultural stability and diversity. *Trends in Cognitive Sciences*, *8*(1), 40-46.

Sperber, D. & Wilson, D. (1995). *Relevance: Communication and Cognition (2nd edition).* Blackwell.

Sperber, D. (1997). Intuitive and reflective beliefs. *Mind and Language, 12*(1), 67-83.

Sperber, D. (2000). Metarepresentations in an evolutionary perspective. In D. Sperber (Ed.), *Metarepresentations: A Multidisciplinary Perspective*. Oxford University Press.

Sperber, D. (2001). An evolutionary perspective on testimony and argumentation. *Philosophical Topics*, *29*, 401-413.

Spiers, H. J., & Maguire, E. A. (2006). Thoughts, behaviour, and brain dynamics during navigation in the real world. *Neuroimage*, *31*(4), 1826-1840.

Spreng, R. N., Mar, R. A., & Kim, A. S. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. *Journal of Cognitive Neuroscience*, *21*(3), 489-510. Squire, L. R. (1992). Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychological review*, *99*(2), 195.

Squire, L.R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory, 82*, 171-177.

Squire, L. R. (1992a). Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychological review*, *99*(2), 195-231.

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

Squire, L.R. & Zola-Morgan, S. (1988). Memory: brain systems and behavior. *Trends in Neurosciences, 11*(4), 170-175.

Stevens, J. R., & Hauser, M. D. (2004). Why be nice? Psychological constraints on the evolution of cooperation. *Trends in cognitive sciences*, *8*(2), 60-65.

Stich, S., & Nichols, S. (1993). Folk psychology: Simulation or tacit theory? *Philosophical Issues*, *3*, 225-270.

Suddendorf, T. & Corballis, M. C. (1997). Mental time travel and the evolution of the human mind. *Genetic, Social & General Psychology Monographs*, *123*(2), 133-168.

Suddendorf, T., & Corballis, M. C. (2007). The evolution of foresight: what is mental time travel, and is it unique to humans?. *Behavioral and brain sciences*, *30*(3), 299-313.

Suddendorf, T., Addis, D. R. & Corballis, M. C. (2009). Mental time travel and the shaping of the human mind. *Philosophical Transactions of the Royal Society B*, *364*, 1317-1324.

Szpunar, K. K. & McDermott, K. B. (2008). Episodic future though and its relation to remembering: evidence from ratings of subjective experience. *Consciousness and Cognition*, *17*, 330-334.

Szpunar, K. K., Spreng, R. N., & Schacter, D. L. (2014). A taxonomy of prospection: introducing an organizational framework for future-oriented cognition. *Proceedings of the National Academy of Sciences*, *111*(52), 18414-18421.

Szpunar, K. K. (2010). Episodic future thought: An emerging concept. *Perspectives on Psychological Science*, *5*(2), 142-162.

Téglás, E., Girotto, V., Gonzalez, M., & Bonatti, L. L. (2007). Intuitions of probabilities shape expectations about the future at 12 months and beyond. *Proceedings of the National Academy of Sciences*, *104*(48), 19156-19159.

Templer, V. L. & Hampton, R. R. (2013). Episodic memory in nonhuman animals. *Current Biology*, *23*(17), R801-R806.

Teroni, F. (2014). The epistemological disunity of memory. In: A. Reboul (ed.), *Mind, Values and Metaphysics: Philosophical Papers Dedicated to Kevin Mulligan – vol. 2* (pp. 183-202). Springer.

Terrier, N., Bernard, S., Mercier, H. & Clément, F. (2016). Visual access trumps gender in 3- and 4-year-old children's endorsement of testimony. *Journal of Experimental Child Psychology*, *146*, 223-230.

Tooby, J. & Cosmides, L. (1992). Cognitive adaptations for social exchange. In: J. Barkow, L. Cosmides & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture*. Oxford University Press.

Tulving, E., Hayman, C.A.G. & Macdonald, C.A. (1991). Long-lasting perceptual priming and semantic learning in amnesia: A case experiment. *Journal of Experimental Psychology: Learning, Memory and Cognition, 17*(4), 595-617.

Tulving, E. (1972) Episodic and semantic memory. In: E. Tulving & W. Donaldson (eds.), *Organization of memory*, (pp. 381–402). Academic Press.

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

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie canadienne*, *26*(1), 1.

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

Tulving, E. (2002b). Chronesthesia: Conscious awareness of subjective time. In D. T. Stuss & R. T. Knight (Eds.), *Principles of frontal lobe function* (pp. 311-325). Oxford University Press.

Tulving, E. (2005) Episodic memory and autonoesis: uniquely human? In: H.S. Terrace & J. Metcalfe (eds.), *The missing link in cognition: Origins of self-selective consciousness*. (pp. 3-56). Oxford University Press.

Turri, J. (2011). Promises to keep: speech acts and the value of reflective knowledge. *Logos & Episteme, 2*(40, 583-590.

Tversky, B. & Marsh, E. J. (2000). Biased retellings of events yield biased memories. *Cognitive Psychology*, *40*, 1-38.

Urmson, J. O. (1967). Memory and imagination. *Mind*, *76*(301), 83-91.

Ünal, E., & Papafragou, A. (2016). Production–comprehension asymmetries and the acquisition of evidential morphology. *Journal of Memory and Language*, *89*, 179-199.

Van Hoeck, N., Ma, N., Ampe, L., Baetens, K., Vandekerckhove, M., & Van Overwalle, F. (2012). Counterfactual thinking: an fMRI study on changing the past for a better future. *Social Cognitive and Affective Neuroscience*, *8*(5), 556-564.

Van Hoeck, N., Watson, P. D., & Barbey, A. K. (2015). Cognitive neuroscience of human counterfactual reasoning. *Frontiers in human neuroscience*, *9*, 420.

von Hippel, W. & Trivers, R. (2011). The evolution and psychology of self-deception. *Behavioral and Brain Sciences, 34*, 1-56.

Vullioud, C., Clément, F., Scott-Phillips, T., & Mercier, H. (2017). Confidence as an expression of commitment: Why misplaced expressions of confidence backfire. *Evolution and Human Behavior*, *38*(1), 9-17.

Waldum, E. R., & Sahakyan, L. (2012). Putting congeniality effects into context: Investigating the role of context in attitude memory using multiple paradigms. *Journal of Memory and Language*, *66*(4), 717-730.

Weiss, M. (1997). Bereavment, commemoration and collective identity in contemporary Israeli society. *Anthropological Quarterly,* 91-101.

Werning, M., & Cheng, S. (2017). Taxonomy and unity of memory. In S. Bernecker & K. Michaelian (eds.). *The Routledge Handbook of Philosophy of Memory* (pp. 7-21). Routledge.

Wheeler, M.E., Petersen, S.E. & Buckner, R.L. (2000). Memory's echo: Vivid remembering reactivates sensory-specific cortex. *Proceedings of the National Academy of Sciences, 97*(20), 11125-11129.

Whitcombe, E. L., & Robinson, E. J. (2000). Children's decisions about what to believe and their ability to report the source of their belief. *Cognitive Development*, *15*(3), 329-346.

Wilson, D. S., Wilczynski, C., Wells, A., & Weiser, L. (2000). Gossip and other aspects of language as group-level adaptations. In: C. Heyes & L. Huber (eds.). *The Evolution of Cognition.* (pp.347-365) MIT Press.

Wimmer, H., Hogrefe, G. J., & Perner, J. (1988). Children's understanding of informational access as source of knowledge. *Child development*, 386-396.

Winter, J. (2001). The memory boom in contemporary historical studies. *Raritan*, *21*(1), 52.

Wright, D. B., Memon, A., Skagerberg, E. M. & Gabbert, F. (2009). When eyewitnesses talk. *Current Directions in Psychological Science*, *18*(3), 174-178.

Wu, J., Balliet, D. & Van Lange, P.A.M. (2015). When does gossip promote generosity? Indirect reciprocity under the shadow of the future. *Social Psychological and Personality Science, 6*(8), 923-930.

Wu, J., Balliet, D. & Van Lange, P.A.M. (2016a). Reputation, gossip, and human cooperation. *Social and Personality Psychology Compass, 10*(6), 350-364.

Wu, J., Balliet, D., & Van Lange, P. A. (2016b). Gossip versus punishment: The efficiency of reputation to promote and maintain cooperation. *Scientific reports*, *6*, 23919.

Xu, F. & Tenenbaum, J. (2007). Sensitivity to sampling in Bayesian word learning. *Developmental Science, 10*, 288–297.

Zeman, A., Dewar, M., & Della Sala, S. (2015). Lives without imagery: congenital aphantasia. *Corext, 73*, 378-380.

ACKNOWLEDGEMENT

TO EXTERNAL FUNDING AGENCIES

CONTRIBUTING TO PHD DISSERTATIONS

Name of doctoral candidate: *Johannes Mahr*

Title of dissertation: *Remembering in Communication - A novel account of the architecture and function of human episodic memory*

Name of supervisor(s): Gergely Csibra, Dan Sperber

External funding agency: **European Research Council**

Acknowledgement:

This research has received funding from the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP7/2007-2013) under **grant agreement No 249519 – OSTREFCOM** and Horizon 2020 programme (E/R/H20/2014-2020) under **grant agreement No 742231 – PARTNERS.**