Review of Philosophy and Psychology (2020) 11:301–333
https://doi.org/10.1007/s13164-020-00471-z

Predicting the Past from Minimal Traces: Episodic Memory and its Distinction from Imagination and Preservation

Markus Werning[1]

Published online: 23 April 2020

Abstract

The paper develops an account of minimal traces devoid of representational content and exploits an analogy to a predictive processing framework of perception. As perception can be regarded as a prediction of the present on the basis of sparse sensory inputs without any representational content, episodic memory can be conceived of as a "prediction of the past" on the basis of a minimal trace, i.e., an informationally sparse, merely causal link to a previous experience. The resulting notion of episodic memory will be validated as a natural kind distinct from imagination. This trace minimalist view contrasts with two theory camps dominating the philosophical debate on memory. On one side, we face versions of the Causal Theory that hold on to the idea that episodic remembering requires a memory trace that causally links the event of remembering to the event of experience and carries over representational content from the content of experience to the content of remembering. The Causal Theory, however, fails to account for the epistemic generativity of episodic memory and is psychologically and information-theoretically implausible. On the other side, a new camp of simulationists is currently forming up. Motivated by empirical and conceptual deficits of the Causal Theory, they reject not only the necessity of preserving representational content, but also the necessity of a causal link between experience and memory. They argue that remembering is nothing but a peculiar form of imagination, peculiar only in that it has been reliably produced and is directed towards an episode of one's personal past. Albeit sharing their criticism of the Causal Theory and, in particular, rejecting its demand for an intermediary carrier of representational content, the paper argues that a causal connection to experience is still necessary to fulfill even the minimal requirements of past-directedness and reliability.

✉ Markus Werning
markus.werning@rub.de

[1] Department of Philosophy, Ruhr University Bochum, 44780 Bochum, Germany

1 Introduction

The current debate in the philosophy of memory, with reverberations also in psychology and neuroscience, is dominated by two camps. On one side, we face modified versions of the Causal Theory holding on to the idea that remembering requires a memory trace that (i) causally links the event of remembering to the past experience of the remembered event and (ii) carries over representational content from the content of experience to the content of remembering. Depending on whether remembering fully or only partially owes its content to the content of a memory trace, these positions can be labelled strong and weak (content) preservationism.[1] Depending on the amount to which new contents are generated, weak preservationism allies with various forms of generationism (Lackey 2005; Michaelian 2011). On the other side, a new camp of simulationists is currently forming up, spearheaded by Kirk Michaelian (2016) in philosophy and Donna Addis (2018) in psychology. Motivated by empirical and conceptual deficits of the Causal Theory and its modifications, simulationists reject not only the necessity of preserving representational content, but also the necessity of a causal link between experience and remembering. They instead view episodic memory in continuity with imagination. As Michaelian (2016) puts it, "[a] ccording to the simulation theory of memory […], the only factor that distinguishes remembering an episode from merely imagining it is that the relevant representation is produced by a properly functioning construction system […] which aims to simulate an episode from the personal past" (p. 97). Hence, for simulationists, the truth-approximating reliability of its production history in conjunction with the personal past directedness of its content is the only property that distinguishes remembering from psychological phenomena such as future, counterfactual, hypothetical, and fictitious imagination.[2] In this spirit, simulationists often apostrophize episodic memory as being merely a peculiar instance of a psychological capacity of mental time (and world) travel, peculiar only with regard to the direction of the time arrow and the actual or another possible world as target.[3]

Albeit sharing many of the objections against the Causal Theory and, especially, against its demand for an intermediary carrier of representational content, I will argue that a causal connection to experience is still necessary to fulfill even the minimal requirements of past directedness and truth-approximating reliability, accepted even by simulationists. I will develop an account of minimal traces devoid of representational content and exploit an analogy to a predictive processing framework of perception. As perception can be regarded as a prediction of the present on the basis of sparse sensory inputs without any representational content, episodic memory can be conceived of as a "prediction of the past" on the basis of a merely *causal* link to a previous experience. The resulting trace minimalist notion of episodic memory will be validated as identifying a natural kind distinct from imagination.

[1] To be distinguished from Senor's (2007) justification preservationism.

[2] In the epistemological literature, the formulation "being the product of a properly functioning system/process which aims at truth" is usually understood as equivocal to "being the product of an (epistemically) reliable system/process". See section 2 for further discussion of the notion of reliability.

[3] The notion of mental time travel has been made prominent by Suddendorf and Corballis (2007). For an attempt to distinguish episodic memory from mental time travel see Cheng et al. (2016). For a review of the relation between episodic memory and mental time travel see Perrin and Michaelian (2017).

I will proceed along the following lines: In the next section, I will outline what I take to be strong and weak versions of the Causal Theory, updating Martin and Deutscher's (1966) original account by contemporary terminology and with amendments necessary to avoid immediate objections. Central notions like that of truth-approximating reliability will be made precise. In section 3, I will review the purported intuitive, explanatory and taxonomical virtues of the Causal Theory. This will draw our attention to a meta-philosophical conflict between the aim of conceptual analysis and the prospects of aligning our psychological concepts with natural kinds. Section 4 leads over to the criticism of the Causal Theory, highlighting the problem of epistemic generativity and the problems of information overkill, overfitting and redundancy. Recognizing the deficits of both strongly and weakly preservationist versions of the Causal Theory, we will arrive at a logical bifurcation: one route leads to Simulationism, the other to Trace Minimalism. As made explicit in section 5, the Simulation Theory gives up not only the need for an (even partial) preservation of representational content, but also the requirement of a causal link between experience and remembering. Trace Minimalism, in contrast, acknowledges the need for a causal link and postulates minimal traces that are devoid of representational content. Pointing to evidence from neuroscience in favor of a particular causal mechanism that links experience and remembering, Trace Minimalism singles out episodic memory as distinct in kind from cases of imagination, even though imaginations may share a number of commonalities with episodic memories. In section 6, the crucial argument against the Simulation Theory is presented, showing that truth-approximating reliability requires a direct or indirect causal link between the event remembered and the event of remembering. The argument will largely draw on a common-cause principle well accepted in the philosophy of science and due to Reichenbach (1956). A number of potential ways-out for Simulationism will be discussed and rejected as inviable. Section 7, finally outlines a view of episodic memory as a prediction of the past from minimal traces, in analogy to the predictive processing theory of perception. The paper concludes with a plea for Trace Minimalism and a rejection of the Simulation Theory and the Causal Theory including their leading variants. Since Trace Minimalism views episodic memory as a natural kind, it supports a discontinuist position with regard to the relationship between episodic memory and imagination.[4]

2 Memory Traces and the Causal Theory

The notion of a memory trace is a defining part of the Causal Theory of Memory. In their seminal paper "Remembering", Martin and Deutscher (1966, henceforth "M&D") tried to develop criteria for a specific kind of memory, which comes close to what psychologists today call episodic memory in contrast to semantic and procedural memory (Squire 1999). According to M&D, someone who, in this sense, remembers an external (e.g., a car accident) or an internal event (e.g., an itch or a thought) at a particular point in time need not only have perceptually, proprioceptively or introspectively experienced that event in the past, and now mentally represent the event with a sufficiently high degree of accuracy. In addition, M&D also postulate a memory trace

[4] For a review of the continuism-discontinuism debate see Michaelian, Perrin, and Sant'Anna (2020, in press).

as a necessary requirement and identify it with a state or process (i.e., a succession of states) that fulfils the following conditions:

(a) *To remember an event, [persons] must not only represent and have experienced it, but also [their] experience of it must have been operative in producing a state or successive states in [them] finally operative in producing [their] representation. […]*
(c) *The state or set of states produced by the past experience must constitute a structural analogue of the thing remembered, to the extent [they] can accurately represent the thing.* (Martin and Deutscher 1966, p. 173 ff.)

Whereas M&D's first condition (a) postulates a causal link between the experience and the remembering and thus speaks in favor of a *causalist* theory of memory, their last condition (c) makes clear that their theory, commonly referred to as *the* Causal Theory, is *not just* a causalist theory. For, M&D require a memory trace not only to provide a causal link between the experience and the remembering, but also to constitute a "structural analogue" of the event remembered such that the memory trace serves as a *representation* of that event. As we will see shortly, the requirement of being a structural analogue is tantamount to viewing those representations as *compositional*. Postulating memory traces as *carriers* of representational content, the Causal Theory adopts a *transmissionist* stance on content preservation (cf. Michaelian and Robins 2018), instead of settling with a relation of mere logical entailment between the contents of remembering and experience. In the course of this paper, I will use the term "preservation" always in the transmissionist sense.

As was noted by epistemologists in the rise of causalist theories of justification in the 1970ies – i.e. after the publication of M&D's seminal paper – one has to build in a condition of reliability in order for causal processes to lend justification to beliefs (Goldman 1979). In doing so, one ensures that counterexamples of deviant or accidental causation are avoided (Plantinga 1993) and the goal of truth is pursued (Werning 2009). Aiming at a modernized formulation of the Causal Theory, I will furthermore assume an almost universally accepted (weak) version of materialism. With these amendments, it seems fair to characterize the Causal Theory's conception of memory traces – in short Traces^{CT} – by the following conditions:

(CT1) **Internality**. Traces^{CT} are internal states or processes of the remembering person. I.e., they are realized by states or processes of the person's present or past (neuro-)biological system.
(CT2) **Causal Link**. Traces^{CT} constitute a causal link between the remembering of an event and the past experience of the event. I.e., the experiential state is an (at least) partial cause of the Trace^{CT}, and the Trace^{CT} is, in turn, an (at least) partial cause of the state of remembering.
(CT3) **Representational Content**. Traces^{CT} are carriers of representational content. That is, Traces^{CT} are compositional representations such that the content of a complex representation is determined by the contents of its parts and the way these parts are structurally combined.

(CT4) [CT4*]
Strong [Weak*] Content Preservation. A Trace's^{CT} content is fully [partially*] contained in the content of the experience, and the Trace's^{CT} content itself fully [partially*] contains the content of the remembering. What is remembered has been fully [partially*] preserved by the Trace^{CT} from what was experienced.

(CT5) **Reliability**. Traces^{CT} are operated in a truth-approximating reliable way by the person's (neuro-)biological system. That is, the process constituted by the Trace^{CT} or essentially involving the Trace^{CT}, with a probability greater than some threshold probability, results in the content of the remembering being sufficiently close to truth.

The five conditions need some interpretational and terminological clarification. The Internality Condition (CT1) reflects M&D's phrase "the state (or successive set of states) *in [them]*". Internality is to be understood here not in the psychological sense as "internally accessible" to the subject, but as – in the physical sense – internal of the person, i.e., as realized by the person's biological states. M&D explicitly reject the idea that memory traces have to be internally – consciously or cognitively – accessible to the subject.[5]

The Causal Link Condition (CT2) is a rather immediate adaption of the original formulation, where "operative in producing" is clearly referring to a causal relation of sorts.

The Representational Content Condition (CT3) tries to capture M&D's requirement on memory traces to "constitute a structural analogue of the thing remembered". Even though M&D apparently want to avoid any commitment to a particular format of representation (e.g., symbolic, pictorial or simulational), their formulation suggests a certain structure-sensitive imaging or mapping relation between the trace and what is remembered. The property of compositionality – in more technical terms – provides exactly that: a structure-sensitive mapping relation, a so-called homomorphism, between the structure of the representational carriers and the structure of their contents. (Fodor 1998; Hodges 2001; Werning 2004, 2005a; Werning et al. 2012 (eds.)). When I, henceforth, speak of representational contents, I – as a matter of terminological convention – imply that their respective bearers are compositional. This precisely means that the content of a complex representation is a structure-sensitive function of the contents of its representational parts. Representational content is, therefore, to be distinguished from merely informational content. It is the latter, but not the former that, e.g., also the angular bend of a bimetal in a thermostat carries with regard to room temperature, due to a probabilistic and functional co-variation relation between the angular bend and the room temperature (Dretske 1988).

The Strong Content Preservation condition (CT4) expresses the content preservationism that goes along with the Causal Theory and is commonly attributed to it. It is also the most frequently and most heavily criticized feature of the Causal Theory. In fact, Strong Preservation seems to be entirely unrealistic and goes against a rich corpus of psychological evidence. As Schacter and Addis (2007, p. 773) summarize this evidence, remembering "is not a literal reproduction of the past, but rather is a

[5] I leave for further exegesis whether M&D throughout their paper accept Internality.

constructive process in which bits and pieces of information from various sources are pulled together". This is why successor theories have tried to incorporate constructive elements. Causal theories of constructive memory (Michaelian 2011; see also Robins 2016), therefore, weaken preservationism and just hold on to Weak Content Preservation (CT4*), according to which only some fraction of the content of remembering is contained in the content of the experience and carried over by the memory trace. Stored semantic information may also contribute to the content of remembering. However, causal theories of constructive memory still maintain that there have to be memory traces that provide a causal link and carry representational content. I will henceforth use the term "Trace^{CT*}" when I refer to only weakly content preserving memory traces, and "Trace$^{CT(*)}$" to comprise both Traces^{CT} and Traces^{CT*}. Likewise, I will use "CT", "CT*" or "CT$^{(*)}$" to refer to the strongly preservationist version of the Causal Theory, the weakly preservationist version and indiscriminately to both versions, respectively.

The Reliability condition (CT5) gives a probabilistic account of the accuracy driven capability M&D might be interpreted as alluding to with the formulation "to the extent [they] *can* accurately represent the thing". (CT5) tries to link M&D's account to the – also otherwise allied, but historically younger – epistemological tradition of process reliabilism (Goldman 1979, 1986; for an adaption to memory see Michaelian 2011). As mentioned earlier, the Reliability condition (CT5) is introduced to allow memory traces to lend justification to beliefs. Notice that the formulation in (CT5) is gradual in two respects: On the one hand, it allows Traces$^{CT(*)}$ to be less than 100% reliable. On the other hand, it does not require achieving absolute truth, but only truth-approximation.

Generally speaking, a belief-forming process is called (epistemically) reliable just in case it leads to true beliefs with a probability greater than some threshold. In the case of memory traces and as a prerequisite of the Reichenbach argument in section 6, the requirement of reliability can be made precise as follows: A Trace$^{CT(*)}$ has to be operated by the biological system S such that, given the resulting memory state $rem_S(\ulcorner e \urcorner)$, the probability of a hit is greater than a certain threshold probability $\rho_{S,E}$, where E is a type of event. I will use Gödel corners to signify representational content such that $\ulcorner e \urcorner$ refers to the event e. By probability I mean objective probability in a frequentist sense, i.e., the proportion of hits to trials would exceed the threshold in the limit of an infinite repetition of trials. The threshold probability $\rho_{S,E}$ must be strictly greater than the a priori probability of a hit. We can hence formulate the condition on reliable production as follows:

(RelP) **Reliable Production.** The remembering event $rem_S(\ulcorner e \urcorner)$ is the outcome of a reliable process of a system S just in case the following inequality holds:

$$P(T(\ulcorner e \urcorner)|rem_S(\ulcorner e \urcorner)) \geq \rho_{S,E} > P(T(\ulcorner e \urcorner)). \quad (1)$$

Here, $T(\ulcorner e \urcorner)$ signifies the verisimilitude, or sufficient truth-closeness, of the representational content $\ulcorner e \urcorner$. The content of the remembering need not be exactly true, but only sufficiently close to the truth. In other words, what counts as a hit is a matter of exceeding a certain threshold $\vartheta_{S,E}$ of truth-closeness (cf. Niiniluoto 1998). Remembering Angela Merkel wearing a dark red suit at the press conference, even though the suit was in fact light red, might, e.g., still count as a hit. The truth-closeness of a

representational content is a matter of the similarity between the (actual or merely possible) event e in the representational content and some event that actually occurs/occurred in our world. Note that, for the representational content to be sufficiently close to truth, i.e., verisimilar, it is not enough that there is a qualitatively similar event existing in the actual world, but this event also has to be similar with regard to its spatial and temporal properties, that is, when and where it occurred. A similarity measure between events has to take this into account. We will call a similarity measure, that takes spatial, temporal and qualitative properties of events into account, an STQ-similarity measure. We can now define verisimilitude as follows[6]:

(SVer) **Verisimilitude.** Given a representational content ⌜e⌝, with e belonging to an event type E, a truth-closeness threshold $\vartheta_{S,\,E}$ ($0.5 < \vartheta_{S,\,E} \leq 1$), and an STQ-similarity measure $d : E \times E \rightarrow [0, 1]$, the satisfaction condition for calling the representational content verisimilar, i.e., T(⌜e⌝), is given as follows:

$$T(\ulcorner e \urcorner) \Longleftrightarrow \exists x.Occur_{@}(\mathrm{x}) \,\&\, d(e,x) \geq \vartheta_{S,E}. \tag{2}$$

This is to say, that, with respect to episodic memory, a representation of an event can be said to be sufficiently close to truth just in case some event actually occurs/occurred whose STQ-similarity to the event represented exceeds the verisimilitude threshold. Aside from being dependent on the biological system S and the type of event E, the reliability threshold $\rho_{S,\,E}$ and the verisimilitude threshold $\vartheta_{S,\,E}$ may also be context-dependent, and, in particular, subject to the epistemic standards invoked in an attribution of memory. Needless to say that for intentional modes other than episodic memory (e.g., perception, testimonial beliefs) different reliability and truth-closeness thresholds might apply.

I would like to point out that the presented characterization of Traces$^{CT(*)}$ leaves open a number of issues. First, with regard to the ontological status of a Trace$^{CT(*)}$, it is left open whether it is identified with a discrete state, with a discrete succession of states or with a continuous process. Second, with regard to its realization, the characterization is equally consistent with a local realization, modelled by a classical computational system or a localist connectionist architecture (where single nodes carry representational content), as with a distributed realization (Sutton 1998), e.g., by a non-localist connectionist architecture, where only patterns of distributed activity carry representational content. Third, the characterization neither presupposes, nor excludes that Traces$^{CT(*)}$ are cognitively or consciously accessible. It leaves open whether these states or processes are personal or subpersonal.

The characterization, moreover, makes no decision with regard to the particular representational format of Traces$^{CT(*)}$ other than that they be compositional. Compositionality does, in particular, not imply that Traces$^{CT(*)}$ are representations in

[6] To take care of the intensional character of the verb *remember*, I distinguish between the existential quantifier $\exists$ and the relational predicate $Occur_w(x)$, 'x occurs in the world w' (Zalta 1988). @ signifies the actual world. For an elaborate treatment of the semantics of *remember* see Liefke and Werning (2018, 2019).

a Language-of-Thought (Fodor 1975) – concatenations of content bearing syntactic symbols ("Mentalese words") – or even symbolic in a more abstract sense, as e.g. proposed by Vector Symbolic Architectures (Smolensky 1995; Stewart and Eliasmith 2012), where there still is a correspondence relation between (non-concatenative) parts in the representational carrier and parts in the representational content. Compositional, but non-symbolic representational architectures have been proposed by Werning (2003, 2012). What compositionality, however, requires is some form of categorization. That is, the content-bearing parts in a representational structure are categorial: They represent either particular entities, (proper or fuzzy) sets of entities or higher set-theoretical constructions built thereon (e.g., propositions). The requirement of categorial representations does, however, not constitute a substantial strengthening of the original Causal Theory. For, M&D themselves assumed a structural analogue between memory traces and the content of remembering, which is commonly construed as involving categories.

I should further mention that there is a difference between having a remembering with a certain content, having a mnemonic belief with that content, and having a belief that is justified by a remembering. The difference is roughly analogous to the difference between having a perception with a certain content, having a perceptual belief and having a belief that is justified by a perception. A person who has a perception with a certain content need not form a perceptual belief with that content. Persons might, for example, have reasons to (erroneously) believe that their experience was an illusion or hallucination, or that it was formed in an otherwise unreliable way. Moreover, a perceptual belief, i.e., a belief with a perceptual (visual, auditory, tactile, olfactory, etc.) content need not be based on a perception. It might be the result of a hallucination or illusion. Perceptions can justify perceptual beliefs, but this justification may be defeasible. Defeaters might come into place; former defeaters might be dropped or defeated themselves. Disentangling the epistemological relations between perceptions and perceptual beliefs has been and still is a quite complex endeavor (Lehrer 1992). The epistemological relations between remembering and mnemonic belief turn out to be no less complex and have received increasing attention over the last two decades (Bernecker 2010; Lackey 2005, 2007; Senor 2007). Trying to avoid issues of justification that go beyond questions of content, verisimilitude and reliability, I will here exclusively focus on rememberings and will disregard questions of mnemonic beliefs and their justification whenever possible.

Last, but not least, I would like to recall that the notion of experience as used by M&D and in my characterization of the Causal Theory is not limited to perception, even though perception is often considered as the paradigmatic case. M&D already mention proprioceptive (e.g., itches) and introspective (e.g., thoughts[7]) experiences. Rememberings might, however, also be based on experiences that are agentive ("I remember what it was like to swim across the river"), emotional ("I remember how angry I was when I learnt about the betrayal") or social ("I remember the humiliation in front of my class mates"). Almost no discussion has yet been devoted to empathic experiences as the source of rememberings, as in cases like: "I remember the feeling of sadness my mother had after

[7] Talking about the introspective experience of thoughts is not unproblematic as Werning (2010) points out in the context of compositionality and transparency requirements.

the death of her brother", where the source of my remembering is not simply my former judgement that my mother was sad, but my empathic experience of her sadness. Our brains seem well equipped for empathic experiences as research on emotional mirror neurons, emotional contagion and the vicarious brain (Ferrari and Coudé 2018; Lamm and Tomova 2018; Olsson and Spring 2018) as well as the simulation theory of empathy (de Vignemont and Singer 2006; Goldman 2006) suggests. The question how far the notion of experience may extend to lay the grounds for remembering will come up in section 6 again. I should mention finally, that, in this paper, I will stick to the widely shared presupposition that, in order to ground the remembering of an event, the experience of the event must be veridical and must have been reliably produced itself.[8]

3 The Purported Virtues of the Causal Theory

Even though I will argue against the Causal Theory and reject both its strongly and weakly preservationist versions, CT and CT*, I do believe that the postulation of memory traces fulfilling (CT1) to (CT5) not only mirrors a long tradition in the history of the notion of memory (Robins 2017) and still reflects the mainstream analysis of the concept of memory in philosophy (Bernecker 2010; Robins 2017), but also, indeed, adequately captures our intuitive concept of remembering in the episodic sense.[9] The different versions of the Causal Theory also capture views strongly influential, if not paradigmatic, in contemporary psychology, where episodic memory – consistent with CT(*) – is usually subdivided into the four stages: experience, encoding, consolidation and retrieval. I therefore do not understand the thrust of this paper as anything close to a conceptual analysis or a historical reconstruction of the concept, nor as a description of its current usage in psychology. The intention of this paper, rather, is the search for natural kinds in the human (and non-human) mind and, in particular, the question of whether episodic memory qualifies as such a natural kind. This may well lead to a revision of our concept of episodic memory.

Even though I disapprove of the Causal Theory(*), I still consider it worthwhile asking why it has been – and still is – so appealing to many. The answer, I suppose, has to do with the frequently shared impression that the Causal Theory(*) aggregates quite some – actual or apparent – intuitive, explanatory and taxonomical support. In the remainder of this section, I will briefly, but not exhaustively go through some of this support – not to defend the Causal Theory(*), but to better understand the motivations behind it.

[8] It would still be worthwhile to ask how to deal with rememberings concerning the content of non-veridical experiences such as dreams, hallucinations and illusions. How can a difference between remembering and misremembering be made with regard to non-veridical experiential content in cases such as *After a night full of dreams, the person remembers [misremembers] a pink elephant chasing after a flying horse* (which is not equivalent to *The person remembers [misremembers] having dreamt of a pink elephant chasing a flying horse*).

[9] For uses of the verb *remember* that do not refer to episodic memory see Werning and Cheng (2017).

3.1 Intuitive and Explanatory Support

The Causal Theory's$^{(*)}$ intuitive support stems from a number of thought experiments. These thought experiments typically assume that, in the past, a person perceived an event and the person now indeed has a true or verisimilar, conscious or at least cognitively accessible representation of the event. However, if one of the following scenarios happens to be the case, we refuse to attribute episodic memory to the person: (i) the person did not form an enduring, not even partial representation of the event on the basis of the genuine perceptual representation, or the person did form it, but it got destroyed (e.g. through a stroke). Still, the person relearns what she had experienced, typically through the testimony of another person, so that the person has a true or verisimilar representation of the event now. (ii) The person indeed formed a lasting internal representation of the event on the basis of genuine perception, but this representation does not cause the current representation of the event. (iii) Same as (ii), except that, this time, the lasting representation indeed causes the conscious or cognitively accessed representation, however, in a non-reliable, perhaps only accidental way. (iv) Similar as in (i), but the person took a record of the event using an external device (e.g., a notebook or drawing) when perceiving it. The person loses her internal representation of the event, but now uses her external representation to form a new internal representation of the event. In contrast to (i), in this scenario, there is a closed causal link between the person's experience and remembering. There is an enduring, though external, content preserving representation, and the whole causal chain might even be reliable.

In the scenarios (i)-(iii), most people would deny that the person remembers the event. Scenario (iv), which basically just challenges the Internality condition, is more controversial. Adherents of an extended cognition view (e.g., Clark and Chalmers 1998) have argued that here the person can justly be said to remember the event.

The Causal Theory$^{(*)}$ may also be evaluated against a number of other explanatory desiderata. Whether it succeeds or not, I would like to leave open here. First, one may ask whether the Causal Theory$^{(*)}$ can, indeed, provide an explanation of the positive justificatory status of remembering. Remembering is regarded as a source of knowledge. The epistemic justification episodic memory provides to beliefs about the past could be traced back via the memory trace to the epistemic justification provided by the experience of the remembered event. One might appeal to the conditions of Causal Link, Representational Content, Strong or Weak Content Preservation and Reliability to secure the truth-conduciveness of episodic memory.[10] The condition of Internality might, in addition, locate and assign the epistemic responsibility for their beliefs to the subjects themselves. A closely related question is whether the Causal Theory$^{(*)}$ might raise the odds for an explanation of the privileged epistemic status of episodic remembering: Unlike

[10] As discussed earlier, the importance of the Reliability condition (CT5), which is often left out in classical formulations of the Causal Theory, must not be overlooked. Whether a causal theory without a reliability requirement would render justification seems questionable to me.

other non-perceptual beliefs, beliefs that are based on episodic remembering are usually assumed to have eye-witness status.

Second, the Causal Link condition might eventually be invoked to account for the reference relation of memories, possibly building on a causal theory of perceptual reference (Raftopoulos and Muller 2006). What determines to which event a particular remembering refers, especially, when the subject has experienced several rather similar events in the past? However, a precise account of how such an explanation would look like is still to be developed.

An even more ambitious goal, finally, would be to explain the quasi-experiential character of episodic memory that Tulving (2005) characterized as a "reliving or re-experiencing" of the past event and that goes along with a certain phenomenal quasi-transparency of remembering as described by Cheng et al. (2016). So far, the only contribution the Causal Theory$^{(*)}$ can claim to make to such an explanation is that memory traces are supposed to carry over representational content from the experience to the remembering.

3.2 A Tool in Taxonomy

Aside from the apparent intuitive attraction and the prospective – or just starry-eyed – explanatory amenities of the Causal Theory$^{(*)}$, the appeal to Traces$^{CT(*)}$ is grounded in the desire to taxonomize episodic memory in distinction to other kinds of mental states. For one, the requirement of a Trace$^{CT(*)}$ would distinguish episodic memory from other forms of declarative memory such as (the various kinds) of semantic memory.[11] Moreover, the appeal to Traces$^{CT(*)}$ would draw a line between episodic memory and imagination. Remembering and imagination might indeed share a number of processes, e.g., some processes involved in the simulation of scenarios. Certain forms of imagination might even draw on some contents of previous experiences – when you, e.g., imagine your next zoo visit, you might largely draw on the experiences you made during previous visits of a zoo. However, according to the Causal Theory$^{(*)}$, remembering your last zoo visit would still be a different kind of state because, only in remembering, your current representation is causally linked to a past experience of an event in a (weakly or strongly) content-preserving way through a reliable process, viz. a Trace$^{CT(*)}$, such that it is this event you remember.

If one were to succeed in pinning down memory traces to an underlying uniform causal mechanism, then episodic memory could indeed be justly called a natural kind. According to the most widely held view on natural kinds, the homeostatic property cluster view (HPC; Boyd 1991; for an application to psychology see Machery 2009), entities should be clustered together, in science, in a way (i) such that the inductive and explanatory potential of theories making reference to those clusters is optimized and (ii) such that this inductive and explanatory potential rests on uniform causal mechanisms underlying each cluster. This leads to the following definition (Cheng and Werning 2016, p. 1358):

(NatK) **Natural Kind.** A class C of entities is a natural kind if and only if there is a large set of properties that subserve relevant inductive and explanatory

[11] The question whether memory, in general, is a natural kind has been addressed before by Michaelian (2010) who argues for a negative answer.

purposes such that C is the maximal class whose members are likely to share these properties because of some uniform causal mechanism.

In the context of their Sequence Analysis of episodic memory, Cheng and Werning (2016) have pointed out, that neuroscience indeed provides evidence for a uniform causal mechanism that links experience and remembering in episodic memory, the so-called replay mechanisms in the hippocampus. For a full account with the relevant scientific evidence, I defer to their original paper, and only provide a short summary here (for illustration see Fig. 1):

The hippocampal replay mechanism generates a record of the sequential succession of experienced events, through alignment, consolidation and reconstruction. Referring to a temporal compression mechanism that is based on theta phase precession and can be observed among hippocampal place cells (Dragoi and Buzsáki 2006; O'Keefe and Recce 1993), Cheng and Werning argue that the hippocampus provides a mechanism

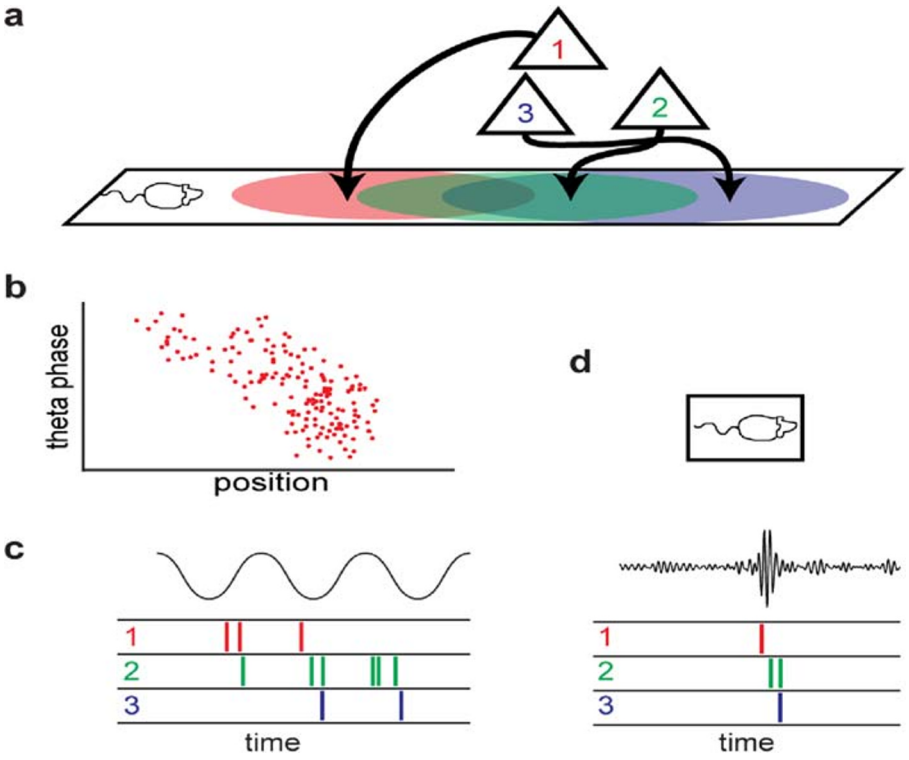

Fig. 1 Schematic illustration of neural activity of hippocampal neurons. (**a**): As the animal explores the linear track, place cells (1,2,3) fire spikes when the animal is located in a circumscribed region in space, the place field (indicated by three ellipses). (**b**)**:** In addition, the spiking of place cells is modulated by the phase of the theta oscillation. Each dot marks the theta phase and position of the animal when neuron 1 is firing a spike as the animal runs from left to right. The correlation between phase and position associated with spikes is known as theta phase precession. (**c**)**:** When spiking of a group of place cells is analyzed within one cycle of the theta oscillation (wavy line at the top), temporal sequences emerge across neurons (theta sequences). (**d**)**:** During the offline state, sharp waves/ripples occur in the local field potential (middle of wavy line at the top, filtered between 150 and 250 Hz) and place cells are reactivated in a sequence that is related to the theta sequences (adapted from Cheng and Werning 2016)

that aligns event sequences in the experiential base with sequences of events in the mnemonic content. For this purpose, patterns of neural activity in the neocortex that realize the experience of individual events are linked to place cells in the hippocampus via the entorhinal cortex (Robinson et al. 2017; Tsao et al. 2018). A temporal compression mechanism generates a record of the sequential succession of the events in the experienced episode at a timescale much shorter (<100 ms) than that of the original sequence of events (≈ 4–5 s). This compression enables consolidation through synaptic plasticity (Azizi et al. 2013). In the offline state, hippocampal replay of the consolidated sequence records and subsequent propagation into the neo-cortex constitute a reconstruction process that produces a mnemonic simulation to represent the past episode (Buzsaki 1989; Gupta et al. 2012; McClelland et al. 1995).

Interventions in the hippocampal replay mechanism indicate that the mnemonic simulations are indeed causally grounded in experiences. Evidence comes from studies on the correlation of neural sequences with short wave ripples and their role for memory and learning in human and non-human animals as well as from computational models (Axmacher et al. 2008; Cheng and Frank 2008; Eschenko et al. 2008). The disruption of short wave ripples impacts long-term memory (Girardeau et al. 2009).

Based on this evidence, the replay mechanism may be hypothesized as constituting an underlying neural mechanism for memory traces. Cheng and Werning leave open, however, whether memory traces, so construed, carry representational content, and consequently, whether representational content is preserved – weakly or strongly – by the memory. The view they have exposed as the Sequence Analysis is consistent with (CT3) and (CT4) as well as with their negation. We will come back to the replay mechanism in section 7.

Cheng and Werning's argument to the conclusion that episodic memories make up a natural kind comes down to the claim that the class of episodic memories is the maximal class of psychological processes that are subserved by the neural mechanism of hippocampal replay, comprising the above described processes of alignment, consolidation, and reconstruction. It remains an open empirical question whether hippocampal replay mechanisms of the same general type can establish a causal link to perceptual experiences of all sense modalities; moreover, to proprioceptive, introspective, agentive, emotional and social experiences; and perhaps even to empathic and vicarious experiences. An answer to this question would finally allow us to determine the maximal class corresponding to the natural kind of episodic memories.[12]

[12] In reply to Andonovski (2018), I would like to stress that the Sequence Analysis was never meant to be a conceptual analysis of remembering, not even a conjunction of necessary and together sufficient properties. It rather expresses a number of properties to specify a prototype of episodic memory that may serve as guidance for the identification of a uniform underlying causal mechanisms. Having once identified that mechanism as one of hippocampal replay, the homeostatic property cluster of episodic memories – corresponding to the natural kind – will amount to the maximal class of phenomena that are subserved by that mechanism. Members of that class may diverge from the prototype captured by the Sequence Analysis. These cases may then be justly called "deficient" cases of episodic memory (e.g., when no longer reliable as in certain patients with dementia). Alluding to an analogy from chemistry, the natural kind of metals – whose underlying uniform causal mechanism is spelled out in terms of Fermi levels in electronic band structure theory – also comprises "deficient" cases such as bismuth (Bi), which differs from prototypical metals, e.g., by very low electric conductivity at 300 K in its pure form.

4 Criticism of the Causal Theory(*)

The Causal Theory has been criticized on various grounds. Some criticisms only target its strong version CT, others also its weak version CT*. As examples one can mention the widely discussed psychological evidence against a strong content preservationism, arguments in favor of a constructivist view on episodic memory (for review: Michaelian 2016; Michaelian and Robins 2018; Roediger and DeSoto 2015), and proposals of memory traces that dispense with a classical view of representational content (Hutto and Peeters 2018; Perrin 2018). Adding to those, I would like to briefly discuss two further objections: (i) CT and CT* contradict the apparent epistemic generativity of episodic memory. They are thus phenomenologically and epistemologically questionable. (ii) There are also information-theoretic reasons against preservationism: Strong preservation would exceed biologically plausible storage capacities and, in the light of the frequent probabilistic correlations in our world, would be massively redundant. Preserving the content of experience can also be rather invaluable for future predictions due to an overfitting effect.

4.1 The Epistemic Generativity of Episodic Memory

An intuitively plausible, perhaps even characteristic feature of episodic memories is that they are epistemically generative, rather than epistemically preservative (Lackey 2005; Mahr and Csibra 2018; Werning and Cheng 2018). An episodic memory may generate new or extra epistemic justification for beliefs where this justification is not itself based on any prior beliefs or other doxastic states (i.e., internal representations with compositional contents).[13] In this respect, episodic memories are analogous to perceptions, which – as is widely held – are not (necessarily) beliefs or doxastic states either. In many cases perceptions ground beliefs. Sometimes, however, for example in cases of known perceptual illusions, perceptual content conflicts with what the subject believes. Moreover, in cases such as the Müller-Lyer illusion, the conflict between perceptual content and belief content is persistent, if not unresolvable. To be epistemically generative, episodic memories must, therefore, not be grounded in beliefs or other doxastic states – such as a memory trace with representational content. In this respect, episodic memories differ from so-called semantic memories, which preserve their justificatory status within the doxastic system.

The epistemic generativity of episodic memory can be illustrated by the following contrast: Case 1: A 49-year-old, who dwelled at the Brandenburg Gate in Berlin on November 9, 1989, is asked about what happened during the Fall of the Berlin Wall. He holds no prior beliefs about whether people were dancing on the wall. In an attempt to remember the event, he explores the situation by generating a mnemonic simulation (Cheng and Werning 2016; Cheng et al. 2016). The simulated scenario provides a justification for him to form the belief that people were dancing on the wall. Case 2: A 20-year-old is asked the same question. She is also justified to believe that people were

[13] Unlike the notion of a belief, which is often assumed to imply a high subjective probability, cognitive accessibility and linguistic expressibility, the notion of a doxastic state is intended to avoid any commitment to those features and to simply refer to internal representations with compositional contents. Traces$^{CT(*)}$, hence, count as doxastic states.

dancing on the wall, but on the basis of recalling what she read about the Fall of the Wall in a school book. In her case no new belief is formed, no extra justification is gained.

The second case is usually explained in terms of the person cognitively accessing a doxastic state. The two cases are phenomenologically so different, though, so that if cognitively accessing a doxastic state explains the second case, it is unlikely to explain the first case equally well. It seems implausible that the person in Case 1 simply accesses a doxastic state he has been entertaining the last thirty years. This would disregard the inquisitive nature of his attempt to remember.

Also in everyday cases, for instance, when I ask myself whether I have turned off the coffee machine at home, constructing a scenario of the past event in episodic memory does not feel like simply retrieving a doxastic state representing me switching off the machine. In the scenario construction, doxastic states might play a role – representing what I usually do after breakfast, where my coffee machine is placed in the kitchen, etc. Still, when I finally succeed in episodically remembering me switching off the machine, this does not feel like the result of some propositional inference. Tulving (1985) plausibly coined the view that episodic memory is more like "re-living or re-experiencing" the past event than like withdrawing an existing doxastic state.

The epistemic generativity of episodic memory strikes me as an argument not only against CT, but also against CT*. If we trust our phenomenology, we may conclude that, whereas the dancing or, respectively, the switching-off are contents of the episodic memories, there was no doxastic state that compositionally represented the dancing or the switching-off, not even partially. Hence, there was also no $\text{Trace}^{\text{CT}(*)}$. As a consequence, not only CT4, but also CT4* seem doubtful.

4.2 Information Overkill, Overfitting and Redundancy

In theoretical neuroscience, it has been relatively well established, that storing all the perceptual contents that CT would require us to preserve in order to be able to remember all the events we are, in fact, able to remember would exceed the informational capacity of our brains. It would lead to an information overkill.

Recently, Richards and Frankland (2017) have, moreover, argued that preserving large amounts of information ("persistence") from a past experience may even have negative effects on one's cognitive performance in the future. Loss of experiential information ("transience") may, in fact, have positive effects. Their argument is based on the overfitting problem: If one fits a past sequence of data too closely, it will become less likely in a noisy world that resulting projections of the data into the future will be close to truth and guide actions in a successful way. Their modelling results can be illustrated by the following real world example: Assume you want to find your favorite restaurant at Hackescher Markt in Berlin – name unknown. The last time you went to Hackescher Markt and visited the restaurant was thirteen years ago. In the meantime, many things have changed: there are many new shops, bars and restaurants, the tram stop was moved, the building of your favorite restaurant was renovated, etc. If your memory contained all past information of this kind from your last visit, you will probably have more problems finding the restaurant, than if you had only remembered the coarse location, say, relative to some street intersection. Richards and Frankland summarize their conclusion as follows: "Transience […] enhances flexibility, by reducing the influence of

outdated information on memory-guided decision-making, and prevents overfitting to specific past events, thereby promoting generalization" (p. 1071).

What seems equally important is that having stored all the contents one remembers in a memory trace would also be excessively redundant. The world we live in is massively loaded with statistical correlations: animals with feathers are very likely to have beaks; when it rains there is a high probability that the sky is cloudy, the air is humid and the streets are wet; when two people fight each other, they are likely to be angry at one another, etc. These are only a few of the simpler correlations between features of objects and between types of events. In the light of those correlations it seems utterly unnecessary, indeed redundant, to store each of the mutually correlated features and events in a memory trace in order to be able, in remembering, to generate a sufficiently verisimilar simulation of a past episode in a sufficiently reliable way.

Moreover, many correlations involve features that are cognitively subcategorial, i.e., they are not categorized by concepts cognitively available to the ordinary human (or non-human) subject. Cognitively subcategorial features and correlations among them may eventually be uncovered, but usually only in a scientific way using refined measurement techniques and an advanced mathematical apparatus. Studying the phenomenon of color constancy, vision scientists have, e.g., discovered numerous correlations between light reflectancies and the retinal stereo-projections of edges, on the one hand, and perceived colors and shapes, on the other (Palmer 1999, for review). These correlations also involve the reflectance properties of the environment, incident light spectra, body postures etc. An object may, e.g., be perceived as a uniformly red cube even though the actual light reflectancies are very complex as are the geometrical patterns projected on the subject's retina. Subjects typically are completely unaware of those complexities because the features involved are, by and large, cognitively subcategorial. Our brains, however, process these subcategorial features and combine them with statistical correlations they have extracted on various levels of abstraction from numerous encounters with stimuli. Once these correlations have been learned – more, precisely: once the synaptic weights among neurons have been adjusted – only a few bits and pieces of sensory information at a time suffice to generate a representation or, better, a simulation of the subject's present environment.

These insights have paved the way for the predictive processing theory of perception (Friston and Kiebel 2009; Hohwy 2013). Here, constructing a simulation of the present is viewed mainly as a top-down, predictive process that iteratively employs the learned statistical correlations on and between various levels of abstraction. The continuously incoming sensory inputs are sparse and merely serve as error signals to be fed back into the prediction process and leading to an update of the simulation. The design principle of the whole system is error minimization. We cannot go into the details of the predictive processing theory of perception here. But what we can say here and now is that, if the theory is right, systems of this kind succeed in reliably generating verisimilar representations of the subject's present. By analogy we might be inclined to infer: If such a system is able to reliably generate verisimilar representations of an episode in the subject's present on the basis of sparse, mostly subcategorial *online* – i.e., sensory – information, it should, in principle, also be able to reliably generate verisimilar representations of episodes in the subject's past on the basis of sparse, potentially non-categorial *offline* information, provided that this information is appropriately linked to the past episode. In both cases, the learned statistical correlations do much of the work. I will explore this line of argument further in section 7.

For now, we can take record of the intermediate conclusion that CT's strongly preservationist requirement of Traces^{CT} would not only challenge the limits of our brain's information capacity and would be even potentially harmful for future predictions and memory-guided action. It would also be excessively redundant given the huge number of statistical correlations in the world. Episodic memory – in the conception of CT – would amount to a massive waste of biologically very expensive resources such that it would very likely not have persisted in evolution.

However, the plausibility of only weakly preservationist Traces^{CT*} can be drawn into question, too. Whereas, weak content preservationism is less affected by the argument from information overkill and overfitting, (CT3) still characterizes Traces^{CT*} as carriers of compositional – and therefore, categorial – content. If the argument from the analogy between the predictive processing account of perception and a – still to be developed – predictive processing account of episodic memory can gain some plausibility, the requirement of carrying over categorial and compositional content from perception to memory loses much of its appeal. Moreover, to the extent that the content of episodic memory is qualitatively and structurally similar to that of experience, subcategorial information and its interplay with learnt statistical correlations might matter no less than the categories that are captured by concepts cognitively available to us and to be embedded in compositional structures. As I will further argue below, storing categorial content is probably superfluous altogether. For these reasons, one should consider rejecting two central tenets of the Causal Theory$^{(*)}$, namely (CT3) and (CT4)/(CT4*).

5 Simulationism Vs Trace Minimalism

If we accept the criticism of the Causal Theory, both in its strong and weak version, and no longer maintain that representational content must be carried over from perception to remembering, we arrive at a logical bifurcation point. The rejection of (CT3) and (CT4)/(CT4*) may lead one to give up the need for memory traces *tout court*, that is, even if they are only characterized by (CT1), (CT2), and (CT5). Accordingly, remembering an event would not have to be causally linked to the perception of the event by an internal process that is operated in a truth-approximating reliable way. This is the route Simulationism takes. According to this view episodic memory consists in nothing but an, in effect, reliable simulation of an episode of one's personal past. The alternative route is Trace Minimalism: the position I will defend in this paper. It holds on to (CT1), (CT2), and (CT5) by developing a notion of a minimal trace: an internal process causally linking experience and remembering and operated by the system in a truth-approximating reliable way. Minimal traces need not bear any representational content and, *a fortiori*, do not preserve it from perception to remembering, not even partially.

5.1 Simulationism

The following citations illustrate how radically Simulationism advocated in Michaelian's book *Mental Time Travel* dispenses with the Causal Theory. Simulationism gives up the idea of memory traces as a requirement on remembering completely and rejects all of the principles (CT1) to (CT5):

> *The simulation theory of memory implies that memory does not have a privileged status relative to other forms of imagination: episodic memory is distinguished from other forms of episodic imagination only by its specific temporal orientation. […T]reating remembering as imagining the past requires abandoning the requirement of a causal connection [between remembering and the past experience] altogether […].* (Michaelian 2016, p. 57)

> *According to the simulation theory of memory defended here, the only factor that distinguishes remembering an episode from merely imagining it is that the relevant representation is produced by a properly functioning episodic construction system […] which aims to simulate an episode from the personal past.* (Michaelian 2016, p. 97)

> *[T]he simulation theory allow[s] that one can in principle remember an entire episode that one did not experience – as long as the relevant representation is of an event belonging to one's personal past, and as long as it is produced by a properly functioning episodic construction system […].* (Michaelian 2016, p. 118)

Simulationism draws its support, for one part, from a slippery slope argument with respect to the constructive character of episodic memory. Since strong content preservationism seems untenable in the light of psychological evidence for constructive elements in episodic memory (Roediger and DeSoto 2015), weak content preservationists already allow that sometimes remembering integrates some non-experiential content. Strong constructivists (Michaelian, p. 103 f., cites Bartlett 1932) further claim that remembering always builds on some non-experiential content. "What [Simulationism] suggests", says Michaelian (2016, p. 104), is that these positions do not go "quite far enough" and concludes that remembering need not be linked to experience at all, neither content-wise nor causally.

This line of thought goes hand in hand with a second class of arguments, based on neuroscientific evidence: As functional neuro-imaging data (from fMRI) suggest, activity in brain regions during episodic remembering largely (though not completely) overlaps with activity in brain regions during episodic imagining (Addis 2018, for review). Moreover, as argued within the scene construction approach (Hassabis and Maguire 2007; Mullally and Maguire 2014), studies, e.g., on patients with hippocampal damage indicate that the hippocampus is not only involved in episodic memory, but also in other episodic construction processes, such as imagining the future or imagining fictitious experiences as well as spatial navigation.

To compensate for the threat of unreliability, finally, Simulationism invokes a metacognitive capacity of source monitoring originally proposed by Johnson et al. (1993), and since supported and linked to underlying brain processes by a number functional neuroimaging studies involving fMRI, PET, EEG and TMS, studies with healthy and brain damaged subjects, as well as modelling studies (for review: Mitchell and Johnson 2009). The problem the source monitoring framework (SMF) addresses is how a particular mental experience – in the broad sense, including episodic memory, episodic imaginations, episodic thought, etc. – is attributed, by the subject, to a source category – such as perception, imagination, memory, dreaming, belief, testimony, etc.,

or even a highly specific source such as "Joe said it" or "I read it in a blue book".[14] SMF starts from the assumption that the content of a mental experience is usually complex and binds together features ranging from perceptual information, spatial and temporal details, semantic information (gist, category membership, associated items), and emotional information to records of the cognitive operations engaged (e.g., imagining or carrying out a mathematical calculation). To provide evidence about the source of a particular mental experience, the elements as well as the characteristics of the mental experience's content, according to Johnson and colleagues, are being related to existing bodies of belief. The characteristics of the content, ranging from vivid to obscure, from detailed to coarse-grained, from intense to feeble, from concrete to abstract, from emotional to emotionless, can be related to average differences in the kind of features that characterize the various sources. Source attribution, according to SMF, may also involve retrieving additional information, discovering and noting relations, extended reasoning as well as other heuristics. The core idea of source monitoring according to SMF, thus, is that the attribution of a particular mental experience to a source – whether the mental experience is based on perception, imagination, dreaming, testimony or other cognitive sources – is a matter of establishing coherence between the elements and characteristics of the experience's content and existing bodies of belief.

The position of Simulationism can be summarized by a number of positive and negative claims about (episodic) remembering:

(SM1) **Simulation**. Remembering is a process of episodic simulation, and, as such, indistinguishable in kind from (other) processes of episodic imagination.

(SM2) **Past-directedness.** In remembering, the simulation represents an episode of one's personal past.

(SM3) **Reliability.** An event of remembering is the outcome of a reliable truth-approximating process of the system in question. Or in other words, the relevant representation is produced by a properly functioning episodic construction system when the system's aim is to simulate an episode from the personal past.

(SM4) **Source Monitoring.** The reliability of the relevant representation is achieved by source monitoring. A sufficiently high degree of coherence between the content of the representation and existing bodies of belief warrants a sufficient degree of reliability.

(SM5) **No Causal link.** In remembering, the simulation of an episode of one's personal past need not be causally linked to an experience of that episode.

(SM6) **No Experiential Base.** It is, a fortiori possible to remember an event that the subject has never experienced at all, as long as the event belongs the subject's personal past.

Michaelian's (2016) claim that episodic memories are subserved by the same type of mechanisms as are imaginations – mechanisms of simulation in an episodic construction system – has a stark consequence: Given that a natural kind – in the HPC sense

[14] SMF rivals with the dual process model, which postulates distinct processes of familiarity and recollection in recognition memory (Eichenbaum et al. 2007; Sauvage et al. 2008).

captured by our definition above (NatK)[15] – comprises the maximal class of entities subserved by a uniform causal mechanism, episodic memories would turn out not to be a natural kind. The relevant set would also have to include episodic imaginations (at least) in order to be maximal. This is true, even though Michaelian proposes to conceptually distinguish episodic memories from imaginations, on the basis of the content criterion of pastness and the epistemic criterion of reliability. However, a mechanism – according to the standard view – is individuated solely by its components, their activities and their regular interactions with each other (Craver 2009; Machamer et al. 2000), and not in terms of representational content or the aim to approximate truth.[16] We can summarize that consequence as follows:

(SM7) **No Natural Kind.** Even though episodic memory can be conceptually distinguished from cases of (mere) episodic imagination – on the basis of the content criterion of pastness and the epistemic criterion of reliability – episodic memory does not constitute a natural kind. That is, the set of episodic memories is a proper subset of the class whose elements share the properties of episodic simulation due to some uniform causal mechanism. The set of episodic memories is not maximal because that class also contains episodic imaginations.

The three main strands of argument for Simulationism are not immune to criticism. The slippery slope argument does, indeed, have some plausibility when it comes to the question whether there have to be memory traces that carry over representational content from experience to remembering. We certainly see a slopy gradient from strongly preservationist to strongly constructivist positions with regard to the amount of representational content to be carried over. The gradient is also slippery insofar as there is apparently no reasonable way to argue where to draw a line between a sufficiently rich amount of representational content and an insufficient amount. So why not give up the requirement of representational content carriers altogether? However, the slippery slope argument is rather unconvincing when applied to the Causal Link condition (CT2). None of the above positions has ever drawn into question the necessity of a causal connection between experience and remembering. As we will see below, the principal reason to postulate a causal link comes from the reliability constraint and is to provide a sufficiently strong probabilistic dependency relation between the event of remembering and the event remembered.

The power of the second strand of argument is also limited. The observation that processes of episodic memory are (partially) co-located, in the brain, with other processes such as future and fictitious imagination does not imply that episodic memory is subserved by the very same types of processes as those forms of episodic imagination. The hippocampal replay account, which lays the empirical ground for the Sequence Analysis of episodic memory, postulates an alignment process between the sequence of events represented during experience in the neocortex and the sequence to be encoded in the hippocampus during consolidation. I do not see how such an alignment process would be a genuine part of episodic imagination. Moreover, the fact that damage to the hippocampus

[15] I will not discuss the question whether episodic memory is a natural kind with regard to any other notion of a natural kind. For, the HPC notion is undoubtedly the most dominant notion of a natural kind in the philosophy of science.

[16] Michaelian (2016) also invokes phenomenological differences to discriminate episodic memory from other forms of imagination. However, phenomenological differences do not make a difference in natural kind, unless one identifies a difference in the underlying mechanisms.

not only affects episodic memory, but also the ability to episodically imagine future or fictitious events, might as well be explained by the hypothesis that many cases of episodic imagination draw on, or even essentially involve episodic memories. When you are asked to imagine your next Christmas Eve celebration, it is often a good strategy to build this imagination on your episodic memories of previous Christmas Eve celebrations. The sequence of events at previous Christmas Eves, ceteris paribus, is a good predictor of the sequence of events at future Christmas Eves. A similar strategy might even help in cases of fictitious imagination, e.g., when it comes to reassembling fragments of various remembered event sequences to generate a new fictitious sequence.

The Source Monitoring Framework, finally, is, of course fully consistent with, at least, the weakly preservationist version of the Causal Theory and certainly with Trace Minimalism. Also here source monitoring is a plausible means to reevaluate the reliability of a memory.

5.2 Trace Minimalism

A way to avoid the strong and the weak version of the Causal Theory, CT and CT*, without succumbing to Simulationism, would be to develop a notion of a minimal trace. A minimal trace is nothing but an internal process that causally links experience and remembering and is operated by the system in a truth-approximating reliable way. Minimal traces need not bear any representational content and, *a fortiori*, need not preserve it from experience to remembering, not even partially. To locate Trace Minimalism in our logical geography, we can say that a process is a minimal trace, Trace™, if and only if it fulfills (CT1), (CT2), and (CT5), and at the same time negates (CT3) and (CT4)/(CT4*). In contrast to Simulationism, Trace Minimalism opposes (SM5), (SM6), and (SM7). Having already presented the arguments against the Causal Theory (CT and CT*), I will further defend Trace Minimalism against Simulationism by arguing that Simulationism is likely to fail with regard to its own criterion of reliability if it rejects the necessity of a causal link between remembering and experience. More precisely, I will argue that (SM3) is in conflict with (SM5) and cannot be rescued by (SM4). Exploring an analogy to the predictive processing theory of perception, I will further show that, by combining minimal traces with learnt statistical correlations, simulations of past episodes can be constructed to represent past episodes in a reliable verisimilar way.

6 No Reliable Production without Causal Connection

Simulationism implies two claims, namely (SM3) and (SM5), that, as I will argue, do not fit together or, at least, cause serious troubles for Simulationism. My argument is conditional on a widely accepted assumption about the relation between causal and probabilistic dependencies that is expressed by what has become known as Reichenbach's Common Cause Principle (Arntzenius 2010; Reichenbach 1956)[17]:

[17] Reichenbach's Common Cause Principle (RCC) is not completely uncontroversial in the philosophy of science, and it seems outright plausible only for non-quantum-mechanical events. It is an open question whether the phenomenon of quantum entanglement can be made consistent with RCC, perhaps by assuming that two entangled events fail to be distinct and metaphysically independent from each other. Challenges also come from situations where order seems to emerge from chaos. RCC is closely related to the Causal Markov Condition. In our context, where we neither face genuine quantum effects nor chaos-theoretical phenomena, I see no objection against presupposing RCC.

(RCC) **Reichenbach's Common Cause Principle.** Let a and b be two distinct and metaphysically independent events and let P be an objective probability function (i.e. chance). Then the following holds: If a and b are probabilistically correlated with one another such that

$$P(a\&b) > P(a)\cdot P(b), \tag{3}$$

then one of three cases is true:

(a) a is a partial cause of b
(b) b is a partial cause of a
(c) There is a common cause c such that c is a partial cause of a and a partial cause of b.

RCC, in other words, states that, if the objective probability of two events a and b occurring conjointly (but not necessarily simultaneously) is greater than the product of the objective probabilities of each of the events a and b, then there is a direct or, through a common cause, indirect causal connection between the events a and b. Cases where a and b are not distinct events – e.g., when one is a constituent of the other – or where a and b are not metaphysically independent (e.g. the buying and the selling of a book) are exempt. Due to the definition of conditional probability, the inequality (3) is logically equivalent to (4):

$$P(\mathrm{a}|b) > P(a). \tag{4}$$

Let us now look at the simulationist Reliability requirement (SM3). Whenever somebody remembers an event e, the remembering $rem_S(\ulcorner e \urcorner)$ should be the outcome of a reliable, truth-approximating process of the relevant system S. If we apply the criterion of Reliable Production (RelP) and assume a probability threshold $\rho_{S,\,E}$ we get the following probabilistic condition:

$$P(T(\ulcorner e \urcorner)|rem_S(\ulcorner e \urcorner)) \geq \rho_{S,E} > P(T(\ulcorner e \urcorner)). \tag{5}$$

We can now insert the definition of verisimilitude (Ver) with respect to some truth-closeness threshold $\vartheta_{S,\,E}$ and some STQ-similarity measure d. This results in the following inequality:

$$P\big(\exists x.Occur_{@}(\mathrm{x}) \wedge d(e,x) \geq \vartheta_{S,E}|rem_S(\ulcorner e \urcorner)\big) > P\big(\exists x.Occur_{@}(\mathrm{x}) \wedge d(e,x) \geq \vartheta_{S,E}\big). \tag{6}$$

Let us now identify e' with some in fact occurrent (past) event to which the remembered event e is sufficiently spatially, temporally and qualitatively similar. That is:

$$e' := \varepsilon x.\big(Occur_{@}(\mathrm{x}) \wedge d(e,x) \geq \vartheta_{S,E}\big). \tag{7}$$

We finally arrive at the following condition for Simulationism[18]:

$$P\left(e'|rem_S(\ulcorner e \urcorner)\right) > P\left(e'\right). \tag{8}$$

This condition instantiates the antecedent of Reichenbach's Common Cause Principle (RCC), i.e., (3), respectively, its equivalent (4). It leads us to the following conclusion: Some in fact occurrent event e' (in the past), sufficiently similar to the remembered event e in spatial, temporal and qualitative respects, is either a partial cause of the remembering of e, or is caused by a common cause c (even further in the past) that also partially caused the remembering of e. The question now is how this can be true given the simulationist's denial of the need for a causal link between experience and remembering as expressed by the claim (SM5). We have to distinguish four cases:

6.1 Direct and Extended Perception

It is the two-step causal connection between the event and perception and between perception and remembering that the Strong and Weak Causal Theory as well as Trace Minimalism regularly choose in order to warrant the truth-approximating reliability of episodic memories. There is a natural causal connection between the event perceived and the event of perceiving. A Trace^{CT}, Trace^{CT*}, or, respectively Trace™ subsequently establishes the causal connection between the event of perceiving and the act of remembering. What is debatable here is what counts as a perception of an event. A more liberal position will allow for extended perception, via optical (e.g., microscopes) or acoustical (e.g., stethoscopes) devices, video and audio (online) transmissions or perhaps even (offline) recordings, etc. Conservative positions might be more restrictive here and may exclude certain technical devices. However, for simulationists, the direct event-perception-remembering connection is not universally available, because they reject the need for a causal connection between perception and remembering (SM5) and even allow for memories without an experiential base (SM6).

6.2 Common Cause Scenarios

The common cause scenario is a logically possible way to establish the probabilistic dependence condition (8). The remembering of an event, $rem_S(\ulcorner e \urcorner)$, and its sufficiently similar counterpart event e' would have to have the same common cause c. To exceed the reliability threshold $\rho_{S,E}$ in a statistically very noisy world like ours, c would have to be very tightly connected to e', probabilistically. This is typically the case when c and e' are parts of a larger event f belonging to an event type F. f might, e.g., be a door-opening event that consists of me hearing my doorbell ring and then turning the doorknob. That I later remember turning the doorknob, might in fact have been caused by the doorbell ringing via my perception of it and a Trace^{CT*} or Trace™ linking it to my remembering, rather than by my turning the doorknob and the experience thereof.

[18] Hilbert's ε-operator in $\varepsilon x.\ \varphi(x)$ reads "some x such that $\varphi(x)$" and is defined by the equivalence $\varphi(\varepsilon x.\ \varphi(x)) \leftrightarrow \exists\, x.\ \varphi(x)$. Strictly speaking, the inference from (6) to (8) involved a type-shift from propositions to events in the argument of the probability function. This can be neglected here, though.

However, this is just the situation where semantic information or statistical correlations are filled in to generate the mnemonic representation of a complex event. This situation, thus, is one that weak preservationists and trace minimalists would allow for anyway and does not provide a good starting point for simulationists.

6.3 Testimony and Vicarious Memory

A causal connection, that establishes the probabilistic dependence condition (8) between the remembering of an event and its sufficiently similar counterpart event in the world, need not necessarily take a route fully internal to the subject (I neglect the bit before the waves etc. emitted from the event hit our senses). As I remarked in section 6.1., if there are certain intermediary technical devices such as cameras and microphones, we may still speak of perception. The more interesting case is testimony where the causal route goes via another person (or multiple persons) and the information about the event to be remembered is conveyed by a verbal report.[19] This is the case Michaelian (2016) apparently wants to allow for.

We can, at least gradually, distinguish between verbal reports that (i) only generate a sparse, verbatim representation in the listener when comprehended and (ii) those that generate a rich, "as if you were there" experience in the listener due to the vividness, detailedness, concreteness, intensity and emotionality of the linguistically conveyed content and of the way the report is told by the speaker. In situations of the first kind, the source monitoring framework, adopted by Michaelian (SM4), predicts that the listener, when recalling the reported event at a later time, will be able to make a clear difference in the sources of his/her memories due to the characteristics of their contents: whether the source was a verbal report of the event or a direct experience thereof. The speaker will thus be able to make a clear distinction between episodic and non-episodic memories. The underlying causal mechanisms will also, for sure, be very different. I would be surprised if anybody would be willing to drop the distinction between episodic memories, with an experiential base, and non-episodic memories, lacking such a base, despite the clear differences the subject is able to make between sources in situations of the first kind.

Situations of the second kind are more promising. Contemporary theories of semantic comprehension indeed assume that understanding a linguistic utterance often involves the construction of a sensorimotorically and emotionally grounded simulation of the utterance's semantic content (Barsalou 2005; Pulvermüller and Fadiga 2010; Werning 2012). If the verbal report has a high degree of vividness, detailedness, concreteness, intensity, and emotionality, the simulations generated during comprehension may indeed be very much like experiencing the event in person. Let's call them narrational vicarious experiences. Research on the sensorimotor and emotional groundedness of linguistic understanding (cited above) indeed informs us that the neural processes, here, largely overlap with actual perceptual, agentive and emotional processes. When the reported event is recalled at a later time, source monitoring processes might, hence, be unable to differentiate a narrational from a live source, due to the great similarity in the characteristics of the remembered content. Memories

[19] Drawings or paintings painting, sometimes, may also count as testimony and may be dealt with in analogy to verbal reports.

of this kind are called vicarious memories (Pillemer et al. 2015) and may, in fact, be regarded as the key witness for the simulationist rejection of the need for an experiential base (SM6). However, to establish the probabilistic dependence condition (8) between the remembering of an event and its sufficiently similar counterpart event in the world, one would still have to postulate a causal connection, consisting of two steps, first between the event and the narrators' testimony and, second, between the listeners' narrational vicarious experience and their remembering the event. An internal Trace^{CT}, Trace^{CT*}, or Trace™ would be needed for the second step. This trace would not link a live experience, but a narrational vicarious experience to the remembering. The Causal Link condition (CT2) may, hence, be upheld against simulationists by including narrational vicarious experiences in the set of experiences. The class of episodic memories, consequently, would eventually have to be extended to include vicarious memories in order to be maximal and count as a natural kind.

The argument, indeed, has the form of a dilemma: If, on the one hand, one acknowledges the existence of episodic memories that are vicarious,[20] simulationists have to postulate a causal connection between the vicarious episodic memory of the event and a vicarious experience of it, in order to establish a probabilistic dependency between the event of remembering and the event remembered and to thereby warrant Reliability (SM3) in light of RCC. For trace minimalists (and causal theorists), in turn, there is no episodic memory without a causal link to experience (CT2). Hence, trace minimalists will identify the causal connection to the vicarious experience with a memory trace (Trace^{MT}) – due to the Representational Content condition (CT3), this would be more difficult for causal theorists. Consequently, trace minimalists will allow vicarious experiences to play the role of experiences in their account of episodic memory. The unpleasant consequence for simulationists, thus, would be that they have to concede a causal connection that the trace minimalist will immediately identify with a Trace^{MT}.

If one, on the other hand, denies that there are any vicarious episodic memories, simulationists do not have a case. They cannot cite the testimony situation as a case of episodic memory without a causal connection to an experience. Scenarios constructed from testimony would just amount to past-directed imaginations built from semantic information provided by verbal testimony.

6.4 Coherence

The appeal to source monitoring (SM4) might suggest that the reliability of a memory could be achieved on the basis of coherence with other bodies of beliefs,

[20] Proponents of classical causal theories – although this is not entailed in CT$^{(*)}$ – often implicitly assume that only "direct experience" counts as a source of episodic memory. Similar presuppositions can be found in the psychological literature on episodic memory. The notion of vicarious episodic memory would thus be regarded as an oxymoron and rejected a priori. I, however, doubt that there is any sensible way to distinguish "direct" from "indirect" experience. Reliable experience always involves a causal chain to the experienced event. The chain can be simple or complex, the relation to the event can be proximal or distal, synchronic or diachronic, personal or impersonal. When you observed a supernova with a telescope, the observation was technically complex and the supernova took place far away, long ago and is in no reasonable sense part of your biography. Still, you did experience the supernova and may well remember it episodically. What matters in the context of episodic memory is whether the experience was reliably produced and perhaps that it, in a certain sense, had some phenomenal quality or transparency (Metzinger 2003; Werning 2010).

alone, and would not require a causal link to the experience of the event in question. This suggestion, due to space limitation, cannot be discussed here. However, it has been scrutinized by Olsson (2017) and answered negatively. Using the Shogenji measure of coherence (Shogenji 1999), Olsson concludes: "Coherence among memories that are independent in the relevant conditional sense but, taken singly, completely unreliable will not have any effect on the posterior probability of those memories" (p. 317). In other words, coherence of memories – that is, episodic and semantic ones, also including other doxastic states – will not lead to a reliable memory of a particular event unless some of the memories made coherent are already reliable with regard to the event in question, and consequently stand in a causal connection to the event.

7 Predicting the Past from Minimal Traces

Trace Minimalism assumes a causal link between experience and remembering and identifies this link with a hippocampal trace. The theory to be develop here largely relies on the CRISP theory of hippocampal replay (Cheng 2013) and its neuro-philosophical interpretation by the Sequence Analysis (Cheng and Werning 2016). According to this view the hippocampal trace encodes a sequential firing of hippocampal place cells. The involved place cells are partially linked to neocortical activation patterns during experience. The full activation pattern, which gives rise to the experiential content, is only fragmentarily connected to the hippocampal trace through a causal mechanism. In the event of remembering, the hippocampal trace invokes a replay of place cell sequences. When reactivated, the place cells project into the neo-cortex and – together with the information flow across neocortical neurons along synaptic connections – give rise to a certain neuronal activation pattern. This activation pattern in the neocortex amounts to the construction of a scenario and can be regarded as carrying the representational content of remembering. Importantly, the scenario construction also crucially involves semantic information stored in the neocortical synaptic connections. This semantic information, over one's lifetime, has been acquired by extracting statistical correlations among cognitively categorial and subcategorial features from experience. The important point is that the hippocampal trace does not carry representational content in its own right, but its constituent cells can be regarded as pointers to only sparse fragments of neural activation patterns that were realizing compositional representations during experience. In the reconstruction process during remembering an entirely new representation is formed on the basis of those sparse fragments of neural patterns. The representational content of the newly formed neocortical activation patterns may be different from that of the past experience due to the reconstructive underdetermination, but also due to synaptic rewiring among the neocortical neurons in the time period between the original experience and the act of remembering.

Figure 2 schematically illustrates the compositionality of experiential and mnemonic content using the modelling framework of Vector Symbolic Architectures (Smolensky 1995; Plate 2003; Stewart and Eliasmith 2012; Werning 2012). Representations correspond to vectors that are realized by distributed activity patterns in the neural network.

To achieve a compositional representation, category representations are bound to thematic roles by ⊗-multiplication of the corresponding category vectors (e.g., *cat*, *bird*, *chase*) with role vectors (*agent*, *patient*, *event*). The resulting vectors are finally ⊕-summed-up to achieve a compositional representation of a complex scenario.[21]

Even though we have representational content in experience and remembering, the hippocampal memory trace, as conceived here, cannot be said to carry representational content and is thus only a minimal trace. For, from the fact that, in experience as well as in remembering, the neocortical activation patterns as wholes carry representational content, it does not follow that some proper fragments of those activation patterns carry representational content. However, it is only such a fragment that is connected to the sequence of place cells in the hippocampus. Moreover, due to changes in the synaptic connectivity in the neocortex over time, the hippocampal trace cannot be said to be stably linked to a neocortical activation pattern with a specific representational content. Finally, even though representational content is not preserved by the hippocampal trace, it could be shown on highly trained neural networks, that a reduction of information to only 1–10% of the experiential pattern of activity still enables the network to reliably generate verisimilar representations in the reconstruction phase (Bayati et al. 2018; Wiskott (in communication)).

To establish the truth-approximating reliability of episodic memory, Trace Minimalism draws on an analogy between the predictive processing framework in perception (Friston and Kiebel 2009; Hohwy 2013) and scenario construction in remembering (Cheng et al. 2016). According to the predictive processing framework, in perception, the brain, more precisely the neo-cortex, produces a prediction about the present on the basis of sparse sensory information and by combining it, in an iterative way, with learned statistical regularities (i.e. semantic information). The bits of sensory information are stored in ultra-short term memory and might be called the sensory trace. The result of this prediction is the content of perception. The sensory trace is the causal link between the sensory processes (stimulation of the sense organs) and the perception. The sensory trace does not carry any representational content, but just cognitively subcategorial sensory information. The resulting prediction can be regarded as a simulation of the present. The sensory trace is reliable if the system is properly functioning in so far, as the result of the prediction has a high probability of being close to truth (in spite of cases of illusions, inattentional blindness, etc.).

By analogy, one may hypothesize that in remembering the brain produces a prediction about the past on the basis of some bits of non-categorial hippocampal information and by combining them with learned statistical regularities. The bits of hippocampal information are minimal traces and causally linked to a previous experience event. The result of this prediction is the content of

[21] Depending on the type of the Vector Symbolic Architecture, ⊗-multiplication is identified with tensor multiplication or cyclical convolution. The latter has the advantage that a dimensional explosion is avoided, recursivity can be implemented and the vectors of syntactic parts can be recovered by inverse operations. For those reasons, the latter provides a proper structure of symbolic representation (Werning 2012). These representations degrade gracefully, given that content similarity is correlated with a similarity of activation patterns. Non-symbolic, but still compositional neural network models of representation have been proposed using neural synchronization as a binding mechanism (Werning 2003, 2005b, 2005c).

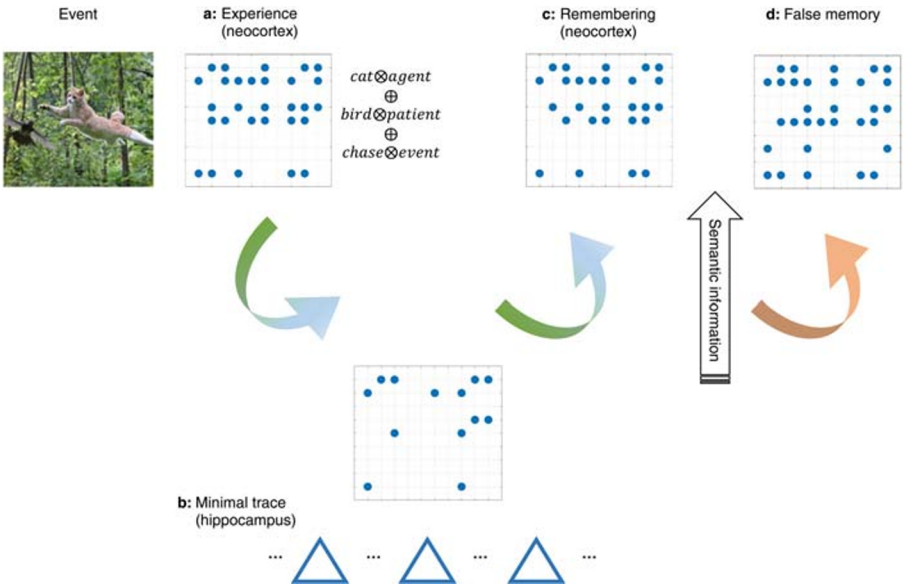

Fig. 2 Schematic view of the interaction of a minimal trace with semantic information**.** The minimal hippocampal memory trace serves as a causal link without representational content between experience and remembering. (**a**)**:** In experience, a distributed implementation of the compositional representation [[*CAT CHASES BIRD*]] is formed, modelled in a Vector Symbolic Architecture (**b**)**:** The minimal trace contains only an informational fragment of the distributed representation present in experience. The fragment does not contain representational content itself. The minimal trace is realized by synaptic connections between place cells (indicated by triangles) in the hippocampus and neocortical neurons (via entorhinal cortex) (**c**)**:** Episodic memory: Due to the interaction with semantic information (stored in the synaptic weights of the neocortex) a verisimilar compositional representation is generated anew. (**d**): False memory: The generated representation has the content [[CAT CHASES BUTTERFLY]]. Albeit matching the informational fragment of the memory trace, a false representation has been generated

remembering. The minimal trace is the causal link between experience and remembering. The minimal trace does not carry any representational content, but just cognitively non-categorial, and sequential hippocampal information. The resulting prediction, i.e., the constructed scenario, can be regarded as a simulation of the past. The memory trace is reliable if properly functioning in so far, as the result of the prediction has a high probability of being close to the truth (in spite of the misinformation effect etc.), given that also the previous experience was reliable.

In other words, constructing simulations of scenarios is a multi-purpose function of the brain and especially the neo-cortex. Its predictive power grounds largely on learnt statistical regularities, encoded in synaptic connections. In perception, the scenario construction machine is fed with subcategorial sensory inputs. The result is a reliable prediction of the present. In remembering, this simulation machine receives non-categorial information from the hippocampus, realized by sparse connections to fragments of neural patterns, once active in experience. The result is a reliable prediction of a past event. No representational content has to be stored; however, a causal link to experience is necessary.

8 Conclusion

The conclusions of this paper can be summarized as follows:

(i) The Causal Theory in its strongly and weakly preservationist variants should be rejected. It is empirically inadequate, from an information-theoretic point of view implausible, and does not account for the epistemic generativity of episodic memory.

(ii) Simulationism turns out to be an unstable position. It tries to hold on to the reliability condition on memory by appealing to a source monitoring framework and at the same time rejects the need for a causal link between experience and remembering. It, however, fails to honor the tight relation that holds between causal and probabilistic dependencies as expressed by Reichenbach's Common Cause Principle. The reliability of an episodic memory amounts to a probabilistic dependency relation between the event of remembering and the event remembered. This probabilistic dependency relation, in turn, requires a direct or indirect causal connection between the remembered event and the episodic memory. Alternative causal routes – including extended perception, a common-cause scenario, and testimony – have been discussed, but would either be unavailable for the simulationist or would require the simulationist to accept a causal link between episodic memory and experience, and thus collapse simulationism into trace minimalism or even stronger positions.

(iii) In the light of empirical evidence and contrary to what simulationism entails, there is good reason to believe that episodic memory will turn out to be a natural kind (in the HPC sense) and as such distinct from imagination. The uniform underlying causal mechanism has been hypothesized to be the mechanism of hippocampal replay. This mechanism can be regarded as a minimal trace causally linking an episodic remembering to a past experience. This does not exclude the possibility that some underlying multi-purpose mechanisms (e.g. constructing simulations) are shared with imaginative processes.

(iv) The notion of experience can, in principle, be widened to also subsume technically extended perceptions and vicarious experiences, along with proprioceptive, introspective, agentive, emotional, social, and perhaps even empathic experiences. Further empirical research is needed to determine the extent to which the various types of experience are causally linked to events of remembering by minimal traces, neurally realized by a general mechanism of hippocampal replay (including alignment, consolidation, and reconstruction). Depending on future empirical findings, the size of the set of episodic memories will be determined as the respective maximal class.

(v) Trace Minimalism rejects the need for memory traces to carry representational content, but demands a causal link between experience and remembering to ensure reliability. It views remembering as a prediction of the past on the basis of sparse, non-categorial information in analogy to the predictive processing theory of perception. Minimal traces constitute such a causal link and thereby provide the physical information to be fed into the brain's scenario construction machinery. Combining the sparse causal-informational trace to a previously experienced event with semantic information on general probabilistic correlations, the scenario construction may reliably yield a verisimilar simulation of the past event.

Acknowledgements This work has been supported by the grants no. 419038924 and no. 419040015 from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) as part of the DFG research group *Constructing Scenarios of the Past* (FOR 2812). I would like to thank Kourken Michaelian and two anonymous reviewers for very helpful comments. I am also grateful to my long-term collaborator Sen Cheng and the members of the research group, in particular to Laurenz Wiskott, Nikolai Axmacher, Annika Dobberke and Anco Peeters, for their input.

Funding Information Open Access funding provided by Projekt DEAL.

References

Addis, D.R. 2018. Are episodic memories special? On the sameness of remembered and imagined event simulation. *Journal of the Royal Society of New Zealand* 48: 64–88.

Andonovski, N. 2018. Is episodic memory a natural kind? A comment on Cheng and Werning's 'What is episodic memory if it is a natural kind' (2016). *Essays in Philosophy* 19: 8–37.

Arntzenius, F. 2010. Reichenbach's common cause principle. In *The Stanford encyclopedia of philosophy*, ed. E.N. Zalta. Metaphysics Research Lab: Stanford University.

Axmacher, N., C.E. Elger, and J. Fell. 2008. Ripples in the medial temporal lobe are relevant for human memory consolidation. *Brain : A Journal of Neurology* 131: 1806–1817.

Azizi, A.H., L. Wiskott, and S. Cheng. 2013. A computational model for preplay in the hippocampus. *Frontiers in Computational Neuroscience* 7: 161. https://doi.org/10.3389/fncom.2013.00161

Barsalou, L.W. 2005. Situated conceptualization. In *Handbook of categorization in cognitive science*, ed. H. Cohen and C. Lefebvre, 619–650. St. Louis: Elsevier.

Bartlett, F.C. 1932. *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.

Bayati, M., Neher, T., Melchior, J., Diba, K., Wiskott, L., & Cheng, S. 2018. Storage fidelity for sequence memory in the hippocampal circuit. PLoS ONE 13(10): e0204685. https://doi.org/10.1371/journal.pone.0204685

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

Boyd, R. 1991. Realism, anti-foundationalism and the enthusiasm for natural kinds. *Philosophical Studies* 61: 127–148.

Buzsaki, G. 1989. Two-stage model of memory trace formation: A role for "noisy" brain states. *Neuroscience* 31: 551–570.

Cheng, S. 2013. The CRISP theory of hippocampal function in episodic memory. *Frontiers in Neural Circuits* 7:88. https://doi.org/10.3389/fncir.2013.00088

Cheng, S., and L.M. Frank. 2008. New experiences enhance coordinated neural activity in the hippocampus. *Neuron* 57: 303–313.

Cheng, S., and M. Werning. 2016. What is episodic memory if it is a natural kind? *Synthese* 193: 1345–1385.

Cheng, S., M. Werning, and T. Suddendorf. 2016. Dissociating memory traces and scenario construction in mental time travel. *Neuroscience and Biobehavioral Reviews* 60: 82–89.

Clark, A., and D. Chalmers. 1998. The extended mind. *Analysis* 58: 7–19.

Craver, C.F. 2009. Mechanisms and natural kinds. *Philosophical Psychology* 22: 575–594.

de Vignemont, F., and T. Singer. 2006. The empathic brain: How, when and why? *Trends in Cognitive Sciences* 10: 435–441.

Dragoi, G., and G. Buzsáki. 2006. Temporal encoding of place sequences by hippocampal cell assemblies. *Neuron* 50: 145–157.

Dretske, F. 1988. *Explaining Behavior: Reasons in a World of Causes*. Cambridge, MA: MIT Press.

Eichenbaum, H., A.P. Yonelinas, and C. Ranganath. 2007. The medial temporal lobe and recognition memory. *Annual Review of Neuroscience* 30: 123–152.
Eschenko, O., W. Ramadan, M. Mölle, J. Born, and S.J. Sara. 2008. Sustained increase in hippocampal sharp-wave ripple activity during slow-wave sleep after learning. *Learning & Memory* 15: 222–228.
Ferrari, P.F., and G. Coudé. 2018. Mirror neurons, embodied emotions, and empathy. In *Neuronal correlates of empathy*, ed. K.Z. Meyza and E. Knapska, 67–77. Amsterdam: Academic Press.
Fodor, J. 1975. *The language of thought*. New York: Crowell.
Fodor, J. (1998). Concepts: Where Cognitive Science Went Wrong. New York, NY: Oxford University Press.
Friston, K., and S. Kiebel. 2009. Predictive coding under the free-energy principle. *Philosophical transactions of the Royal Society B: biological sciences* 384: 1211–1221.
Girardeau, G., K. Benchenane, S.I. Wiener, G. Buzsaki, and M.B. Zugaro. 2009. Selective suppression of hippocampal ripples impairs spatial memory. *Nature Neuroscience* 12: 1222–1223.
Goldman, A.I. 1979. What is justified belief? In *Justification and knowledge*, ed. G. Pappas, 1–25. Dordrecht: Reidel.
Goldman, A.I. 1986. *Epistemology and cognition*. Cambridge: Harvard University Press.
Goldman, A.I. 2006. Simulating minds: The philosophy, psychology, and neuroscience of mindreading. *Oxford: Oxford University Press.*
Gupta, A.S., M.A. van der Meer, D.S. Touretzky, and D.D. Redish. 2012. Segmentation of spatial experience by hippocampal θ sequences. *Nature Neuroscience* 15: 1032–1039.
Hassabis, D., and E.A. Maguire. 2007. Deconstructing episodic memory with construction. *Trends in Cognitive Sciences* 11: 299–306.
Hodges, W. 2001. Formal features of compositionality. *Journal of Logic, Language and Information* 10: 7–28.
Hohwy, J. 2013. *The predictive mind*. Oxford: Oxford University Press.
Hutto, D.D., and A. Peeters. 2018. The roots of remembering: Radically enactive recollecting. In *New directions in the philosophy of memory*, ed. K. Michaelian, D. Debus, and D. Perrin, 97–118. New York: Routledge.
Johnson, M.K., S. Hashtroudi, and D.S. Lindsay. 1993. Source monitoring. *Psychological Bulletin* 114: 3–28.
Lackey, J. 2005. Memory as a generative epistemic source. *Philosophy and Phenomenological Research* 70: 636–658.
Lackey, J. 2007. Why memory really is a generative epistemic source: A reply to Senor. *Philosophy and Phenomenological Research* 74: 209–219.
Lamm, C., and L. Tomova. 2018. The neural bases of empathy in humans. In *Neuronal correlates of empathy*, ed. K.Z. Meyza and E. Knapska, 25–36. Amsterdam: Academic Press.
Lehrer, K. 1992. *Theory of knowledge*. London: Routledge.
Liefke, K., and M. Werning. 2018. Evidence for single-type semantics—An alternative to e/t-based dual-type semantics. *Journal of Semantics* 35: 639–685.
Liefke, K., & Werning, M. (2019). Single-type semantics and depiction reports. In *13th International Tbilisi Symposium on Language, Logic and Computation*. Tbilisi: Tbilisi State University (unpublished).
Machamer, P., L. Darden, and C.F. Craver. 2000. Thinking about mechanisms. *Philosophy of Science* 67:1–25.
Machery, E. 2009. *Doing without concepts*. Oxford: Oxford University Press.
Mahr, J., and G. Csibra. 2018. Why do we remember? The communicative function of episodic memory. *Behavioral and Brain Sciences*: 41. https://doi.org/10.1017/S0140525X17000012.
Martin, C.B., and M. Deutscher. 1966. Remembering. *Philosophical Review* 75: 161–196.
McClelland, J.L., B.L. McNaughton, and R.C. O'Reilly. 1995. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review* 102: 419–457.
Metzinger, T. 2003. Phenomenal transparency and cognitive self-reference. *Phenomenology and the Cognitive Sciences* 2: 353–393.
Michaelian, K. 2010. Is memory a natural kind? *Memory Studies* 4: 170–189.
Michaelian, K. 2011. Generative memory. *Philosophical Psychology* 24: 323–342.
Michaelian, K. 2016. *Mental time travel: Episodic memory and our knowledge of the personal past*. Cambridge, MA: MIT Press.
Michaelian, K., D. Perrin, and A. Sant'Anna. 2020. Continuities and discontinuities between imagination and memory: The view from philosophy. In *The Cambridge handbook of imagination*, ed. A. Abraham. Cambridge: Cambridge University Press (in press).
Michaelian, K., and S. Robins. 2018. Beyond the causal theory? Fifty years after Martin and Deutscher. In *New directions in the philosophy of memory*, ed. K. Michaelian, D. Debus, and D. Perrin, 13–32. New York: Routledge.
Mitchell, K.J., and M.K. Johnson. 2009. Source monitoring 15 years later: What have we learned from fMRI about the neural mechanisms of source memory? *Psychological Bulletin* 135: 638–677.

Mullally, S.L., and E.A. Maguire. 2014. Memory, imagination, and predicting the future. *The Neuroscientist* 20: 220–234.

Niiniluoto, I. 1998. Verisimilitude: The third period. *The British Journal for the Philosophy of Science* 49: 1–29.

O'Keefe, J., and M.L. Recce. 1993. Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* 3: 317–330.

Olsson, A., and V. Spring. 2018. The vicarious brain: Integrating empathy and emotional learning. In *Neuronal correlates of empathy*, ed. K.Z. Meyza and E. Knapska, 7–23. Amsterdam: Academic Press.

Olsson, E.J. 2017. Coherentism. In *The Routledge handbook of philosophy of memory*, ed. S. Bernecker and K. Michaelian, 310–322. New York: Routledge.

Palmer, S. 1999. *Vision science: Photons to phenomenology*. Cambridge, MA: MIT Press.

Perrin, D. 2018. A case for procedural causality in episodic recollection. In *New Directions in the Philosophy of Memory*, ed. K. Michaelian, D. Debus, and D. Perrin, 33–51. New York: Routledge.

Perrin, D., and K. Michaelian. 2017. Memory as mental time travel. In *The Routledge handbook of philosophy of memory*, ed. S. Bernecker and K. Michaelian, 228–239. New York: Routledge.

Pillemer, D.B., K.L. Steiner, K.J. Kuwabara, D.K. Thomsen, and C. Svob. 2015. Vicarious memories. *Consciousness and Cognition* 36: 233–245.

Plantinga, A. 1993. *Warrant and proper function*. Oxford: Oxford University Press.

Plate, T. 2003. *Holographic reduced representations*. Stanford: CSLI Publications.

Pulvermüller, F., and L. Fadiga. 2010. Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews. Neuroscience* 11: 351–360.

Raftopoulos, A., and V. Muller. 2006. Nonconceptual demonstrative reference. *Philosophy and Phenomenological Research* 72: 251–285.

Reichenbach, H. 1956. *The direction of time*. Berkeley: University of Los Angeles Press.

Richards, B.A., and P.W. Frankland. 2017. The persistence and transience of memory. *Neuron* 94: 1071–1084.

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

Robins, S. 2017. Memory traces. In *The Routledge handbook of philosophy of memory*, ed. S. Bernecker and K. Michaelian, 76–87. New York: Routledge.

Robinson, N.T.M., J.B. Priestley, J.W. Rueckemann, A.D. Garcia, V.A. Smeglin, F.A. Marino, and H. Eichenbaum 2017. Medial Entorhinal Cortex Selectively Supports Temporal Coding by Hippocampal Neurons. *Neuron* 94: 677–688.e6.

Roediger, H.L., and K.A. DeSoto. 2015. Reconstructive memory, psychology of. In *International encyclopedia of the Social & Behavioral Sciences*, ed. J. Wright, 2nd ed., 50–55. Oxford: Elsevier.

Sauvage, M.M., N.J. Fortin, C.B. Owens, A.P. Yonelinas, and H. Eichenbaum. 2008. Recognition memory: Opposite effects of hippocampal damage on recollection and familiarity. *Nature Neuroscience* 11: 16–18.

Schacter, D.L., and D.R. Addis. 2007. The cognitive neuroscience of constructive memory: Remembering the past and imagining the future. In: *Philosophical transactions of the Royal Society B: biological sciences*362: 773–786.

Senor, T.D. 2007. Preserving Preservationism: A reply to Lackey. *Philosophy and Phenomenological Research* 74: 199–208.

Shogenji, T. 1999. Is coherence truth conductive? *Analysis* 59: 338–345.

Smolensky, P. 1995. Connectionism, constituency and the language of thought. In *Connectionism*, ed. C. Macdonald and G. Macdonald, 164–198. Cambridge, MA: Blackwell.

Squire, L. 1999. Memory, human neuropsychology. In *The MIT encyclopedia of the cognitive sciences*, ed. R.A. Wilson and F.C. Keil, 520–522. Cambridge, MA: MIT Press.

Stewart, T., and Eliasmith, C. 2012. Compositionality and biologically plausible models. In *The Oxford handbook of compositionality*, ed. M. Werning, W. Hinzen, & E. Machery, 596–615. Oxford: Oxford University Press.

Suddendorf, T., and M.C. Corballis. 2007. The evolution of foresight: What is mental time travel, and is it unique to humans? *Behavioral and Brain Sciences* 30: 299–313.

Sutton, J. 1998. *Philosophy and memory traces: Descartes to connectionism*. Cambridge: Cambridge University Press.

Tsao, A., J. Sugar, L. Lu, C. Wang, J.J. Knierim, M.-B. Moser, and E.I. Moser. 2018. Integrating time from experience in the lateral entorhinal cortex. *Nature* 561: 57–62.

Tulving, E. 1985. Memory and consciousness. *Canadian Journal of Psychology* 26: 1–26.

Tulving, E. (2005). Episodic memory and autonoesis: Uniquely human? In *The missing link in cognition: Origins of self-reflective consciousness*, ed. H. Terrace and J. Metcalfe, 3–56. Oxford: Oxford University Press.

Werning, M. 2003. Synchrony and composition: Toward a cognitive architecture between classicism and connectionism. In *Applications of mathematical logic in philosophy and linguistics*, ed. B. Löwe, W. Malzkorn, and T. Raesch, 261–278. Dordrecht: Kluwer.

Werning, M. 2004. Compositionaltity, context, categories and the indeterminacy of translation. *Erkenntnis* 60: 145–178.

Werning, M. 2005a. Right and wrong reasons for compositionality. In *The compositionality of meaning and content*, ed. M. Werning, E. Machery, and G. Schurz, vol. I, 285–309. Frankfurt: Ontos Verlag.

Werning, M. 2005b. The temporal dimension of thought: Cortical foundations of predicative representation. *Synthese* 146: 203–224.

Werning, M. (2005c). Neuronal Synchronization, Covariation, and Compositional Representation. In *The Compositionality of Meaning and Content*, ed. E. Machery, M. Werning, and G. Schurz, vol. II, 283–312. Frankfurt: Ontos Verlag.

Werning, M. 2009. The evolutionary and social preference for knowledge: How to solve Menon's problem within reliabilism. *Grazer Philosophische Studien* 79: 137–156.

Werning, M. 2010. Descartes discarded? Introspective self-awareness and the problems of transparency and compositionality. *Consciousness and Cognition* 19: 751–761.

Werning, M. 2012. Non-symbolic compositional representation and its neuronal foundation: Towards an emulative semantics. In *The Oxford handbook of compositionality*, ed. M. Werning, W. Hinzen, and E. Machery, 622–654. Oxford: Oxford University Press.

Werning, M., and S. Cheng. 2018. No need for meta-representation: How scenario construction explains the epistemic generativity and privileged epistemic status of episodic memory. *Behavioral and Brain Sciences*. 41. https://doi.org/10.1017/S0140525X17001534.

Werning, M., and S. Cheng. 2017. Taxonomy and unity of memory. In *The Routledge handbook of philosophy of memory*, ed. S. Bernecker and K. Michaelian, 7–20. New York: Routledge.

Werning, M., W. Hinzen, and E. Machery, eds. 2012. *The Oxford handbook of compositionality*. Oxford: Oxford University Press.

Zalta, E.N. 1988. *Intensional logic and the metaphysics of intentionality*. Cambridge, MA: MIT Press.