Philos Stud
DOI 10.1007/s11098-016-0647-x

Representing the past: memory traces and the causal theory of memory

Sarah Robins[1]

Abstract According to the Causal Theory of Memory (CTM), remembering a particular past event requires a causal connection between that event and its subsequent representation in memory, specifically, a connection sustained by a memory trace. The CTM is the default view of memory in contemporary philosophy, but debates persist over what the involved memory traces must be like. Martin and Deutscher (Philos Rev 75:161–196, 1966) argued that the CTM required memory traces to be structural analogues of past events. Bernecker (Memory: A philosophical study. Oxford University Press, Oxford, 2010) and Michaelian (Philos Psychol 24:323–342, 2011), contemporary CTM proponents, reject structural analogues in favor of memory traces as distributed patterns of event features. The proposals are understood as distinct accounts of how memory traces represent past events. But there are two distinct questions one could ask about a trace's representational features. One might ask how memory traces, *qua* mental representations, have their semantic properties. Or, what makes memory traces, *qua* mental representations of *memories*, distinct from other mental representations. Proponents of the CTM, both past and present, have failed to keep these two questions distinct. The result is a serious but unnoticed problem for the CTM in its current form. Distributed memory traces are incompatible with the CTM. Such traces do not provide a way to track the causal history of individual memories, as the CTM requires. If memory traces are distributed patterns of event features, as Bernecker and Michaelian each claim, then the CTM cannot be right.

Keywords Causal theory of memory · Remembering · Memory traces · Mental representation

✉ Sarah Robins
skrobins@ku.edu

[1] Department of Philosophy, University of Kansas, 3090 Wescoe Hall, 1445 Jayhawk Blvd, Lawrence, KS 66045, USA

Published online: 04 March 2016

1 Introduction

According to the Causal Theory of Memory (CTM), remembering a particular past event requires a causal connection between that event and its subsequent representation in memory. There may be many such connections. The CTM aims to privilege just one: the connection sustained by a memory trace. A memory trace is a representation of the past event, providing a link that is informational as well as causal. The CTM is enormously influential—earning the right to be called the default view of memory in contemporary philosophy, if anything does[1]—but debates over the best characterization of its memory traces are ongoing. When Martin and Deutscher (1966) introduced the CTM, they argued that memory traces are structural analogues of past events. Bernecker (2010) and Michaelian (2011), contemporary CTM proponents, reject structural analogues in favor of a view of memory traces as distributed patterns of event features.

These proposals are understood as distinct answers to the following question: how do memory traces represent past events? As posed, the question is ambiguous. It runs together two distinct questions one might ask about a memory trace's representational features. It might be asking how memory traces, *qua* mental representations, have their semantic properties. Or, it might be asking what makes memory traces, *qua* mental representations of *memories*, distinct from other mental representations. The first concerns content determination generally; the second targets memory traces specifically. But proponents of the CTM, both past and present, have failed to keep these two questions distinct. The result is a serious but unnoticed problem for the CTM in its current form. Distributed memory traces are incompatible with the CTM. If memory traces are distributed patterns of event features, as Bernecker and Michaelian each claim, then the CTM cannot be right.

I begin by distinguishing these two representational questions—the Content Determination Question and the Memory Trace Question—in further detail. Next, I turn to CTM as introduced by Martin and Deutscher, arguing that their answer to the Memory Trace Question invokes two features of memory traces: what they represent (trace content) and when they were acquired (trace causal history). The representational and causal-historical features work in tandem to distinguish remembering from other similar mental states. This point is obscured in Martin and Deutscher's own presentation. The claim that memory traces are structural analogues receives more attention, both in their analysis and in responses to it from Bernecker and Michaelian who take such analogues to be implausible. But Martin and Deutscher's appeal to structural analogues is an answer to the Content Determination Question. It is unnecessary for articulating a theory of memory, and so can be removed from the CTM while leaving its unique answer to the Memory Trace Question intact. Finally, I turn to the updated versions of the CTM provided by Bernecker and Michaelian. Their rejection of structural analogues seems

[1] For example, Martin and Deutscher's article is the only entry on memory included in a collection of major epistemology papers of the twentieth century (Bernecker and Dretske 2000). It is also routinely cited as the default view of memory, against which others offer refinements of explore implications (e.g., Shoemaker 1970; Palmer 1978; Deutscher 1989; Bernecker 2010; Debus 2010; Michaelian 2011).

promising initially, but unfortunately, the alternative account of memory traces they each endorse conflicts with the CTM answer to the Memory Trace Question. The distributed network account of memory traces does not provide a way to track the causal history of individual memories. Since these traces cannot be individuated by when they were acquired, they lack one of the two key features the CTM requires. Bernecker, Michaelian, and other memory theorists must choose between their allegiance to the CTM and their commitment to distributed traces. I conclude with a discussion of how views of representing the past should move forward.

2 How do memory traces represent past events?

Memory traces are mental representations of past events, mental states in virtue of which remembering previous experiences is possible. How do memory traces represent past events? The CTM offers a particular answer to this question, distinct from the characterization provided by other theories of memory. For the moment, our interest is not in these answers but the question asked. The intrigue it invites may make it difficult to recognize that more than one thing is being asked. There are at least two questions that can be distinguished:

> *Content Determination Question*: How do memory traces—*qua* mental representations—have their semantic properties?
>
> *Memory Trace Question*: What makes memory traces—*qua* mental representations of memories—distinct form other mental representations?

The *Content Determination Question* is one of general interest to philosophers of mind, psychology, and cognitive science. Mental representations play a prominent (if not defining) role in psychological explanation. All theorists who appeal to representations when explaining the mental capacities of humans and other cognitive creatures owe an answer to the Content Determination Question. In addition to an account of *what* is represented one must also provide an account of *how* this is achieved—in virtue of what do mind/brain states have the semantic properties that they do?[2]

The *Memory Trace Question* is more narrow, asking about the details of only a subset of mental representations. After all, there are many kinds of representational mental states—beliefs, desires, intentions, as well as (possibly) various perceptual and emotional states. The Memory Trace Question is posed to those interested in characterizing memory as one amongst the many on this list. The distinction could be drawn along any number of dimensions. Memory traces could be distinguished by their *location*. Alternatively, they could be distinguished by the *attitude* taken toward a mental content. Memory could be a kind of propositional attitude, perhaps distinguished by a concomitant phenomenological marker like familiarity (Broad 1925), pastness (Russell 1912), or intimacy (James 1890). Memory traces could also be distinguished by their *content*, for example, as representations of events in the

[2] In framing the question this way, I am following Von Eckardt (1996).

past tense (e.g., Locke 1971). And, finally, it could be some other *functional*, *modal*, or *historical* property that helps to define memory traces. The CTM is a view of this last sort, on which memory traces are distinguished at least in part by their causal history. This interpretation of the CTM is the focus of Sect. 4.

Memory theorists will want complementary answers to these two questions. One's account of *what* memory represents may be especially well suited to a particular view of *how* content-structure mappings are assigned, or vice versa. Whatever interrelations one seeks, the questions remain distinct. And it is the former that is of primary importance for articulating a theory of memory. Insofar as a proponent of the CTM, or any other account of memory, cares about questions of mental representation, their interest is in the representational features of memory traces and so in the Memory Trace Question.[3]

An account of what is distinctive about memory representation can be given without an accompanying explanation of how memory's distinctive contents are determined. Of course, the Content Determination Question lingers. But the burden of answering it does not fall to the memory theorist alone; it is shared by all theorists who appeal to mental representations. In fact, a memory theorist has good reason to avoid endorsing any particular view of content determination prematurely, at least insofar as the theory of memory is built to rely on it. Doing so would expose the account to unnecessary risk, should its endorsed theory of content determination fall from favor. It would render the commitment to memory's existence, and to memory traces, susceptible to elimination by factors that should not have a direct influence. Accounts of what memory represents should be driven primarily by considerations about memory. A better approach, it seems, is to provide a characterization of memory's content that leaves the question of how its contents are mapped onto the mind/brain's underlying structure to be determined—a blank that can be filled in by whatever the best theory of content determination turns out to be.

3 The causal theory of memory

Martin and Deutscher's (1966) analysis of remembering identifies two challenges that any version of the CTM must address, regardless of whether one also endorses their proposed solutions. I introduce these challenges and their responses here. Evaluation is withheld until Sect. 4. The analysis focuses on the requirements for remembering a particular past event. It begins with the following general constraint:

> *Causal Condition*: To remember an event, a person must not only represent and have experienced it, but also his experience of it must have been operative

[3] Memory theorists who reject the role of memory traces in an account of memory (e.g., Bennett and Hacker 2003) or who argue for a purely logical account of traces (e.g., Rosen 1975) must answer this question, too, in a modified form that does not presume memory's distinctiveness is derived from the representations in which it traffics. Such views are, however, beyond the scope of the present paper.

> in producing a state or successive states in him finally operative in producing his representation (Martin and Deutscher 1966: 173).[4]

This constraint guards against the possibility of veridical confabulation: cases where one serendipitously reencounters information that was never fully learned or otherwise forgotten.[5] Even with veridical confabulation excluded, there are ways that the connection could be sustained, resulting in an accurate representation of the past event while nonetheless failing to qualify as remembering. Martin and Deutscher consider two such possibilities—*relearning* and *nonmemorial retention*—as challenges compelling further conditions on the CTM.

3.1 Relearning

There are many ways that information from a past event can be retained across time. Experiences can be recorded in memory, but also in books, time capsules, and video recordings, as well as in the minds of others who learn about the event secondhand. These sources often provide helpful cues, prompting a would-be rememberer to recall an event that might otherwise be difficult to bring to mind. But the availability of these sources, together with the frailty of memory, creates the possibility for *relearning*: the reacquisition of forgotten information. Suppose a person discovers a box of her old journals in the attic and reading through them results in thoughts about events she had not considered for decades. Have the journals been a mere prompt, or are they a source for relearning? The Causal Condition alone does not allow for the cases to be distinguished. A further condition must be added to allow prompting while excluding relearning.

One might expect that the desired distinction could be made easily, because the difference between remembering and relearning is transparent to the person in question. But Martin and Deutscher resist appeal to any phenomenological markers to distinguish remembering from relearning. The *feeling* of remembering is defeasible: even the most vivid experiences of seeming to remembering can be inaccurate. People often take themselves to be remembering when they are not, and may disavow veridical representations of their own past experiences.[6] One might also assume that prompting and relearning could be distinguished by appeal to the *amount* of information provided by the external cue. Prompts are often only fragments, whereas relearning requires that all of the information be supplied. But

[4] Martin and Deutscher take the constraints identified in the opening clause of this condition to be widely accepted, if under specified. The causal requirement is their unique contribution. By “operative” they mean that it must play a causal role in bringing about the representation that serves as an instance of remembering.

[5] In this way, motivation for the CTM runs analogously to that given for the more prominent Causal Theory of Perception, where the causal connection between a perceptual object and the perceptual experience is required to rule out cases of veridical hallucination (Grice 1961; Lewis 1980).

[6] While relearning, a person may recognize that the information is coming from an outside source, but even so, may forget the source on subsequent occasions when the event is called to mind. Martin and Deutscher insist this is commonplace, as ‘recollections’ from one’s early childhood illustrate (Martin and Deutscher 1966: 182).

Martin and Deutscher resist this as well. They maintain that remembering is possible even in cases of *strict prompting*, where the prompt contains all of the details the rememberer could possibly provide (Martin and Deutscher 1966: 182).

Excluding cases of relearning, they argue, requires appeal to the causal history of the state involved in producing the representation of the past event. A mental state acquired during relearning accurately represents the past event, but it lacks the right connection to that event. It was formed long after, not during, the event being represented. This distinction forms the basis of Martin and Deutscher's second condition:

> *Operative-In Condition*: In those cases where prompting is operative for the representation, his past experience of the thing represented is operative in producing the state (or successive set of states) in him which is finally operative in producing the representation *in* the circumstances in which he is prompted (Martin and Deutscher 1966: 185).

Remembering and relearning differ because the former ability to produce the representation stays *within the person* in some important (if so far unspecified) sense. In relearning, this ability has been lost and must be reacquired.[7]

3.2 Nonmemorial retention

The analysis of remembering requires further disambiguation. Another cognitive ability—also acquired during the past event—could produce a representation of that event, albeit inadvertently. Since the ability was acquired during the event that one is trying to remember, it cannot be distinguished from remembering by appeal to causal history. Excluding such cases of *nonmemorial retention* requires a way to sort between representations produced by memory and representations produced by other, nonmemorial cognitive processes.

Martin and Deutscher envision a case of suggestibility, where a person is hypnotized and acquires a tendency to suggestion. Whenever this suggestible state is active, the person will endorse any information presented. Suggestibility is problematic when a person is strictly prompted about the hypnosis event. Suppose the person was hypnotized while attending a birthday party. If suggestibility is induced and the person is strictly prompted about the party, then she will accept the information from the prompt. The *Operative-In Condition* allows for such a case. Her representation has the right causal history, but the wrong feature of that historically apt state is operative.[8]

[7] Martin and Deutscher wrestle with this boundary, recognizing that proximity to the body serves as a poor marker for a representation staying *inside* the rememberer. As they note, relearning can occur even when the information never leaves the body and, further, remembering need not involve physical contact between the memory and the body.

[8] Hypnosis has been shown to increase subjects' susceptibility to misinformation (e.g., Laurence and Perry 1983). For a recent study remarkably similar to what Martin and Deutscher envision, see Mendelsohn et al. (2008).

The problem of nonmemorial retention can be raised without appeal to hypnosis. A range of capacities could produce accurate but inadvertent representations of the event during which they were acquired. Suppose our person is a lifelong sufferer of major depression, the onset of which coincided with the birthday party mentioned above. When asked to recall the party, her retelling is colored by a current bout of depression (i.e., it is filled with negative judgments of the event). This may be accurate; she may have had a miserable time. In this condition, however, she is not remembering. Given her depressive state, she would have characterized the party negatively even if it had been pleasant at the time.[9]

Nonmemorial retention challenges the possibility of memory for events during which a person acquires a long-standing disposition that, when active, influences whether or how information is represented. It is therefore tempting to block these cases by ruling out the possibility of ever remembering such events. Martin and Deutscher resist this restriction; a person prone to suggestibility or depression may be able to remember the event when these capacities are inactive. What is needed is a condition that provides "an effective statement of the difference between the memory part of *M* [the previous state] and the suggestible part of *M*" (Martin and Deutscher 1966: 188). Of all the abilities acquired during and retained since a past event, how can the ability to remember be distinguished? Martin and Deutscher propose the following:

> *Structural Analogue Condition*: The state or set of states produced by the past experience must constitute a structural analogue of the thing remembered, to the extent to which he can accurately represent that thing (1966: 191).

The memory part of the mental state can be distinguished from the nonmemorial because it alone is a representation of the event.

4 Memory traces as structural analogues

Martin and Deutscher claim memory traces as an "indispensable part of our idea of memory" (1966: 189), required for meeting the challenges above. What's more, they specify what memory traces must be like—memory traces are structural analogues of the events they represent. In this section, I evaluate Martin and Deutscher's account of memory traces with an eye toward the two representational questions distinguished in Sect. 2. I make two claims. First, their account of traces as structural analogues is an answer to the Content Determination Question, not the Memory Trace Question. An account of content determination is unnecessary and so their final condition can be weakened to a more generic mental representation requirement. Second, I argue that this weakened condition is sufficient for the analysis. It handles the nonmemorial retention challenge, in so doing, makes clearer the CTM's unique answer to the Memory Trace Question.

[9] The depression example comes from Bernecker (2010), although he does not interpret it as an instance of nonmemorial retention.

Martin and Deutscher offer little elaboration on the *Structural Analogue Condition*. They concede an “inevitable recourse to metaphors” (Martin and Deutscher 1966: 189) in any characterization of traces. In fact, Martin and Deutscher do not offer any reason for characterizing traces as structural analogues specifically and it is unclear how the addition of this feature would help to address nonmemorial retention. As soon as it becomes apparent that solving the nonmemorial retention problem will require mental representations, Martin and Deutscher invoke memory traces as structural analogues. This seamless transition—from traces as representations to traces as structural analogues—is indicative of their unreflective commitment to a particular view of mental representation, namely content determination via structural isomorphism. The commitment is not a memory-specific concern. That is, tucked into Martin and Deutscher’s account is an assumption that *all* mental representations should be understood in this way.[10]

Whether this account of mental representation succeeds is, for present purposes, beside the point. The *Structural Analogue Condition* answers the Content Determination Question. It is an account of how traces have their semantic properties; it is not an account of what distinguishes memory traces from other states with semantic properties. The condition goes beyond what is required for answering the Memory Trace Question. It can be avoided by simply removing the effects of this assumption from the final condition, replacing “structural analogue” with “mental representation” as follows:

> *Mental Representation Condition*: The state or set of states produced by the past experience must constitute a *mental representation* of the thing remembered, to the extent to which he can accurately represent that thing.

The condition is now suitably generic. Memory traces are representations of the event. How these (and other) representations are implemented in the mind/brain, as the Content Determination Question asks, can be filled in by whatever theory of content determination turns out to be the right one.

Even in this weakened form, the condition addresses the nonmemorial retention challenge. The condition must distinguish remembering from other ways that a capacity acquired during the past event, like suggestibility or depression, could generate an accurate representation of the past event. The *Mental Representation Condition* does this: memory traces are distinguished because they alone are representations of the event. Although there may be circumstances where memory, suggestibility, and depression produce similar results, the nature of each underlying ability is distinct. Remembering is the ability to produce representations of the event as it was experienced. Suggestion is the ability to endorse the information with which one is presented. And so on. Even if the other abilities are representational, the contents represented will be different.

[10] The assumption is common. Many philosophers and cognitive scientists have endorsed such a view, especially at the time Martin and Deutscher were writing. Structural isomorphism is a view of representation via resemblance. To say that a memory trace, or any representation, is a structural analogue of what it represents is to say that the structure of the representation mirrors the structure of what is represented (see Palmer 1978).

Weakening this final condition also allows us to see more easily how Martin and Deutscher intend to distinguish memory traces from other representational mental states—i.e., how the CTM answers the Memory Trace Question. Memory traces are distinguished by what they represent *and* when they were acquired. Both features are equally important. Although Martin and Deutscher do not talk of memory traces until addressing nonmemorial retention, constraints on traces emerge in their earlier discussion of relearning. Relearning cases are excluded by appeal to causal history, when the mental state was acquired relative to the event being represented. The relearning and nonmemorial challenges are best viewed as complementary. In relearning, the representation is brought about by the right kind of capacity (i.e., a capacity to represent an event), but the state lacks the right causal connection to the event. In nonmemorial retention, the representation has the right causal history, but is not produced by the right kind of capacity. The representational and causal-historical features of memory traces must work in tandem; remembering requires both. According to the CTM, remembering a past event requires a representation that was produced during the event in question, in virtue of the information about the event that is represented.

5 Updating the causal theory of memory

The CTM has persisted relatively unchanged since Martin and Deutscher's analysis, as Michaelian explains:

> There has been relatively little work in recent years on the metaphysics of memory (the nature of memory in general). This is presumably because most philosophers have assumed that something close to Martin and Deutscher's (1966) causal theory of memory is right (2011: 323).

Renewed interest in memory has increased attention to their analysis in recent years. Contemporary philosophers of memory have, by and large, positioned themselves as CTM proponents (e.g., Bernecker 2010; Debus 2010; Michaelian 2011). They understand their main task to be updating the CTM's view of memory traces, replacing structural analogues with an account more hospitable to recent work in memory science. Below I focus on Bernecker and Michaelian, each of whom defends a *distributed* account of memory traces.[11]

5.1 Rejecting structural analogues

Bernecker (2010) and Michaelian (2011) each object to Martin and Deutscher's characterization of memory traces as structural analogues. Although their concerns are distinct, the argumentative strategy is similar—both challenge the account of content determination that supports the Structural Analogue Condition.

[11] I do not include Debus' revised CTM because she does not challenge or replace the structural analogue requirement.

Bernecker takes issue with the account of mental representation on which the Structural Analogue Condition is based, declaring it “wholly implausible” (Bernecker 2010: 134). The account of representation fails because the mirroring it requires cannot be obtained:

> First, to say of something that it has a structure is to say that it consists of parts and that these parts are configured in a certain way. But what are the parts of the world and of mental states which, according to the structuralist view of traces, resemble one another? Second, neither the mind nor the world have a single natural, non-arbitrary structure of elements. Rather the world and the mind have as many different structures as there are ways of describing them—which is to say an infinite number of structures (Bernecker 2010: 134).

In short, memory traces cannot be structural analogues because isomorphism is neither necessary nor sufficient for representing something. There is no inherent structure in virtue of which representations, mental or otherwise, can be guaranteed their content.[12] The objection, broadly construed, poses a challenge for all theories of mental representation that rely on structural isomorphism. Bernecker harnesses this point and directs it at Martin and Deutscher’s account of memory traces, but the point of the criticism remains focused on content determination.

Michaelian’s objection to structural analogues emerges as part of his defense of the CTM from Zemach’s (1983) challenge that memory theorists “dictate to science” by insisting that traces exist and that they have a particular structure instead of leaving these questions open to empirical investigation. Michaelian insists that the existence of traces has been empirically vindicated, claiming that “there has been no suggestion from empirical science that memory might function without memory traces of some sort” (Michaelian 2011: 331). Commitment to the existence of traces *reflects* memory science; it does not dictate to it. But Michaelian endorses Zemach’s criticism of dictating to science regarding what memory traces are like. It is wrong for Martin and Deutscher to prejudge how traces will represent past events. And further, even though it is possible that scientists will discover that memory traces are isomorphic to the events they represent, Michaelian believes that memory science has all but confirmed that memory traces are “holographic” instead.[13] Whether or not this is right, the nature of Michaelian’s objection to Martin and Deutscher is clear. When Michaelian rejects the structural analogue account of traces in favor of the holographic alternative, he is advocating for an alternative account of content determination.

Bernecker and Michaelian’s criticisms of the *Structural Analogue Condition* are consistent with my claim in Sect. 4 that this condition offers an answer to the Content Determination Question. They both claim that memory traces cannot be structural analogues because there is something wrong with the view of mental

[12] For an extended argument on this point, see Heil (1981).

[13] The suggestion of the holographic alternative begins in Zemach, but he is more agnostic than Michaelian about its empirical status. Neither Zemach nor Michaelian elaborates much on the details of this proposal. Zemach says only that the account appeals to“holographic principles challenging the notion of localized codes of the sort used, e.g., in a phonograph record” (Zemach 1983: 33).

representation on which this account of traces relies. The condition either makes implausible demands or fails to reflect the discoveries of our best memory science. Either way, it should be removed from the CTM. I turn now to Bernecker and Michaelian's proposals for its replacement.

5.2 Revising the CTM analysis

Bernecker and Michaelian each defend an updated version of the CTM, framed as an account of the necessary and sufficient conditions for establishing that *S remembers that p*. Bernecker and Michaelian's accounts are distinct, but for the purposes of this paper it is their similarities that are of primary interest. They agree on which features of Martin and Deutscher's view should be retained and, for the features they reject, they agree on both the source and nature of the alternative.

Let's start with what is retained. Martin and Deutscher's analysis began with a few background assumptions about memory. Remembering requires a person to represent a past event that did in fact occur. Bernecker and Michaelian each agree, and build requirements of accurate representation into their own accounts. The heart of Martin and Deutscher's analysis is their commitment to remembering as a causal relation, spelled out across three conditions: the *Causal Condition*, the *Operative-In Condition*, and the *Structural Analogue Condition*. Bernecker and Michaelian retain versions of the first two, albeit in forms modified to suit their respective accounts. That is to say, they agree that memory requires a causal connection between the initial event and its subsequent representation in memory, and further, that the connection must be sustained by an internal, retained state of the rememberer in order to preclude cases of relearning.

Here is Bernecker's:

S's representation at t2 that p is causally connected to S's representation at t1 that p* such that

- S's representation at t1 that p* and S's representation at t2 that p are connected by a persisting memory trace or contiguous series of memory traces,
- the memory trace is at least an inus condition for S's representation at t2 that p. If the memory trace is an independently sufficient condition, it is not preempted by another independently sufficient condition,
- If S hadn't represented at t1 that p* he wouldn't represent at t2 that p (Bernecker 2010: 242).

Here is Michaelian's:ϕ

- There is a causal chain running back from S's current representation of p to an earlier representation of hers, and
- The causal chain goes continuously via a (distributed) memory trace with the content P or something sufficiently close to P (Michaelian 2011: 335).

Bernecker and Michaelian's versions of the CTM also differ from Martin and Deutscher's in two key respects. First, both Bernecker and Michaelian allow for some changes to the memory's content over time. This could happen because the rememberer receives some new information and updates the memory, or because some of the details are forgotten. The content changes cannot be drastic. Bernecker characterizes the old and new contents as *p** and *p*, respectively, and requires an entailment relation between them: "p and p* supervene on the same environmental conditions at t1 or p is entailed by p* (where entailed is understood along the lines of relevance logic)" (Bernecker 2010: 242). Michaelian's trace condition, quoted above, states that the contents must be "sufficiently close." Second, and most importantly, neither Bernecker nor Michaelian offers a direct replacement for Martin and Deutscher's *Structural Analogue Condition*. Instead, Bernecker and Michaelian both believe that the account of traces should come from our best science of memory and, further, that our best memory science characterizes memory traces as *distributed*. A detailed account of these traces, and the problems they present to these updated versions of the CTM, awaits in the next section.

6 Memory traces as distributed patterns

Bernecker and Michaelian both express general support for a revised account of memory traces that derives from memory science. Neither elaborates on the nature of these traces in much detail, and so it is worth quoting in full the minimal characterizations they do provide.

Here is Bernecker's:

> It is a mistake to think of the way information is stored in traces by analogy to impressions on a wax tablet or grooves in a vinyl record. Instead psychologists tell us that information is stored in the strengths of connections between neurons. On this picture traces are distributed rather than local: what can be distinctly remembered need not be held distinctly or independently, since each memory item is spread or 'superposed' across many elements in a neural network (2010: 137).

And Michaelian's:

> There is overwhelming evidence from empirical investigations of memory that memory involves traces of precisely the holographic (or distributed or superpositional) sort mentioned by Zemach (2011: 332).

The key resource here is Sutton's (1998) discussion of memory traces, to which both Bernecker and Michaelian appeal as an expression of the account of distributed traces they have in mind. Sutton defends a "connectionist" account of traces, which "employ superpositional storage: memories are blended, not laid down independently once and for all, and are reconstructed rather than reproduced" (Sutton 1998: 2). From this characterization, we can extract three key features of distributed traces. They are (1) stored superpositionally, (2) blended, and (3) reconstructed. I discuss each of these features in tern.

First, superpositional storage. The basic idea is that memory comprises a network of information, with nodes corresponding to each idea, concept, or event feature a person encounters. The connections between nodes in this network reflect associations between the corresponding ideas. Suppose as an (admittedly oversimplified) example, that some of the nodes in this network are: <party>, <dog>, <cake>, <bicycle>, and <leash>. In such a network we would expect the <dog>–<leash> and <party>–<cake> connections to be strong, especially in comparison with <leash>–<cake>. Memories are stored as a pattern of connections between nodes in the network. The memory of a birthday party, for example, might be represented as a pattern that connects <party>–<cake>–<birthday>–<gifts>–<bicycle>. All memories are stored within this network, and so it follows that memories of events with shared features will have overlapping patterns. All memories of parties will include <party>, and many will also involve <cake>, but they might differ in terms of whether they branch out to include <birthday>, <holiday>, or <retirement>. The more similar the events remembered, the more overlap in their patterns. To say that a memory is stored superpositionally is thus to say that the memory is distributed throughout the network rather than being stored in a distinct local node and that the distributions of similar events will overlap.

The account of superpositional storage makes clear how memories become blended. The repeated overlap of event features explains the differences in connection strength between various nodes. Event features that frequently co-occur will become stronger over time—the <party>–<cake> connection is strong because parties often involve cake. By blending memories in this way, not laying them down “independently once and for all,” as Sutton puts it, the system allows memory to track patterns of association across events. It does so at the expense of retaining the specifics of any particular past event. Even though each memory is encoded as a specific pattern of connections between event features, the patterns do not necessarily have a lasting impact. Similar event patterns blend together in composite representations. The effect of any particular pattern will wash out over time, as the encoding patterns from other events changes the strength of connections across the network. The connection between <party>–<balloons> may be especially strong when a person is younger, for example, but may weaken as he or she grows older and parties become more associated with nice dinners and wine. As the <party>–<wine> connection strengthens, the patterns for memories of balloon-involving parties fade.

Since memories are superpositional and blended, it no longer makes sense to characterize remembering as retrieval or reproduction. Instead, the remembering process is, as Sutton describes, *reconstructive*—the memory is built, at the time of recall, by activating a pattern in the network, cobbling together associated event features to produce a representation of the event one wants to remember. The constructive process is guided by the state of the network at the time of remembering. The pattern created will depend on the associations available. Associations between nodes are determined by frequency and recency (i.e., which event features have co-occurred most often and which have just co-occurred). In remembering, construction begins from a cue, the node(s) associated with what the rememberer wants to recall—<party>, for example. Activation then spreads to associated nodes. The nodes that are most easily accessible from <party> will be those to which it is best connected, either because they were recently activated alongside <party> or because they are often activated alongside <party> (or both).

This reconstruction may not be an accurate representation of the event one is trying to remember. Given the nature of this network, the constructed pattern will privilege common features of parties, even if those features were not involved in the particular party one intends to represent. The tendency to favor recent and frequent information is considered an advantage of this account of memory traces, even if it means that memory's contents can be distorted. In fact, its ability to explain distorted memories is precisely why many memory scientists favor this account of memory traces. Applying connectionist principles to memories for particular past events has proven useful for explaining many features of remembering, including the possibility and predominance of memory errors and the construal of remembering as a constructive process. By distributing memories of particular events across the network, connectionist accounts of traces more easily explain how details of similar events are confused for one another and how activation of these patterns is used to build (rather than simply retrieve) representations of the past (e.g., Wixted et al. 2014).

6.1 The representational question

As I have just shown, Bernecker and Michaelian explain how memory traces represent past events by appeal to a distributed account of memory traces. We are now in a position to ask: is this account an answer to the Content Determination Question or the Memory Trace Question?

At first glance, distributed traces look like an answer to the Content Determination Question. The networks across which these traces are spread are often characterized as "connectionist" and connectionism is an account of content determination. Moreover, it's an account of content-structure mapping often invoked as an alternative to structural isomorphism. It seems plausible, therefore, that contemporary CTM proponents have intentionally selected an account of content determination that opposes the one supporting structural analogues. Connectionism is a rejection of the idea that mental representation requires a tight mirroring between the representation's content and vehicle, as structural isomorphism insists. Instead, connectionist networks distribute a mental representation's content across a sub-symbolic set of units (e.g., Smolensky 1991).

The similarity between connectionist networks and distributed traces, however, is misleading. Distributed traces, as described above and employed in the CTM, are not connectionist networks distributing mental content. While it's true that the memory traces are characterized as distributed rather than local, what is being distributed are *memories*, not *mental contents*. Connectionist networks and distributed traces share certain structural features. Both employ networks of interconnected nodes and distribute content across them, but they differ in the level of content at which this is done.

The difference is best illustrated visually. Figure 1 depicts a connectionist network of mental content. Contents are represented as patterns of connections in this network. *Party* might be represented as the pattern 2-4-6. *Birthday* might be 1-5-8, and *bicycle* 2-5-7. Importantly, the connectionist network is sub-symbolic—its nodes are not meaningful units. The representation of *bicycle* has multiple

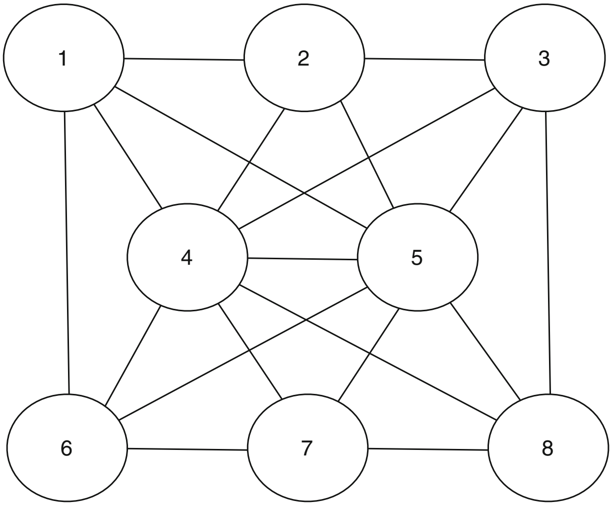

Fig. 1 Connectionist network of mental content

constituents, but lacks a constituent structure. There is no mapping between the meaningful elements of the concept *bicycle* and the nodes in the pattern that represents it. Put another way, the nodes in the network are not semantically evaluable. The ‘2’ node does not have a fixed meaning that it contributes to each representation in which it features. The pattern for *bicycle* may not overlap with the pattern for *tricycle* even though they overlap in content, and the pattern for *bicycle* may overlap with the pattern for *birthday*, as it does in this example, without any content shared between them.

Figure 2 depicts a network of distributed memory traces. In this network, memories are represented as patterns of connections between event features. A recent birthday party for a friend may be represented as the pattern <wine>–<birthday>–<party>–<restaurant>. The pattern <party>–<retirement>–<cake>–<balloons> may represent a recent work event. What is distributed here is the memory—the representation of a particular past event is spread across a set of event features. But there is no further distribution of mental content implied. Each node in the network is semantically evaluable; <party> has a fixed meaning that it contributes to each representational pattern in which it features.

Distributed memory traces are not connectionist networks. They are not, therefore, an answer to the Content Determination Question. Distributed traces are not an account of how the semantic properties of mental representations are secured by the structure of the mind/brain. Instead, distributed traces are an account of how memories in particular are represented—as patterns of event features in a semantically evaluable network. Distributed traces are an answer to the Memory Trace Question. According to the account of distributed traces that Bernecker and Michaelian appeal to in their updated versions of the CTM, memory is organized so as to privilege patterns across events rather than details within any particular event.

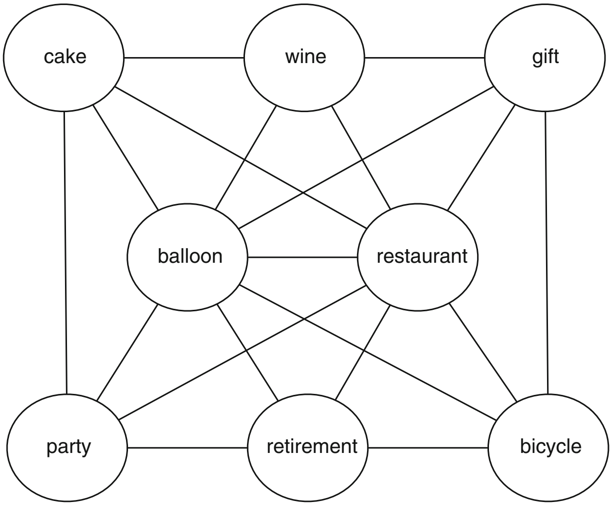

Fig. 2 Network of distributed memory traces

6.2 A problem for distributed traces

The account of memory traces as distributed patterns appears to have several advantages. First, it focuses on the right representational question, offering an answer to the Memory Trace Question rather than the Content Determination Question. This allows Bernecker and Michaelian to avoid Martin and Deutscher's troublesome commitment to structural analogues, a second advantage. What's more, it refreshes the CTM by aligning it with the view of memory traces now favored in memory science. Thus far, the replacement of structural analogues with distributed patterns appears to be a step in the right direction.

Unfortunately, the advantages are only apparent. Bernecker and Michaelian's commitment to memory traces as distributed patterns is in conflict with the CTM. The answer to the Memory Trace Question that such traces provide is different than the answer found in Martin and Deutscher's original analysis. Worse, the two answers are mutually inconsistent. If memory traces are distributed as Bernecker and Michaelian suggest, then the CTM cannot be right.

In Sect. 4 I argued that close attention to Martin and Deutscher's analysis reveals the CTM's unique answer to the Memory Trace Question. The CTM distinguishes memory traces by appeal to two features: what is represented (trace content) and when the representation was acquired (trace causal history). The challenges of nonmemorial retention and relearning, respectively, make clear the importance of each. Distributed traces have the first feature, as one would expect of an account of memory traces—a trace is a representation of the remembered event. It is the second feature that presents the problem. Distributed traces do not have individually distinguishable causal histories. Traces are stored as patterns of event features, but these patterns are superpositional such that similar memories are blended together in overlapping patterns. Individual traces do not leave a lasting, distinctive mark on the network by which their unique causal influence on a subsequent representation could be detected, much less distinguished from the influence of any other distributed pattern.

To see the point, consider how the process of remembering would work for the distributed memory trace network in Fig. 2. Suppose a person forms a memory of a recent party, stored in this network as <birthday>–<party>–<wine>–<restaurant>. Later, she attempts to remember this event by wondering which was the last party she attended. For distributed traces, the process of remembering is reconstructive: the representation is built by compiling the contents available in the nodes that are strongly connected to the node serving as the cue. They will be those whose connections to the cue are the most recent and/or the most frequent. In this example, the remembering process begins with the cue <party>. Because the associations with <wine>, <birthday>, and <restaurant> are the most recent connections to <party>, it is possible to reconstruct the pattern corresponding to this event—i.e., for the representation to be accurate. But the influence of this event as a unique pattern in the network is fleeting. As soon as any of the involved features become part of the pattern for another event—as soon as the would-be rememberer drinks wine at a gallery opening or goes to a restaurant for a work dinner—the connection strengths between <party> and <wine> or <restaurant> will begin to change. Attempts to remember the last party attended after such shifts in the network will result in changes to the representational pattern reconstructed. At this point the event is, in some sense, forgotten. Its pattern continues to be a contributor to the network's connection strengths, but not in such a way that its unique influence remains detectable. In fact, even when that unique pattern was recoverable, the representation produced was not necessarily influenced by it alone. Suppose that the person in question often attends parties where wine is served. The frequent co-occurrence of these features will result in a strong <party>–<wine> connection, such that reconstructed patterns that begin by using either of these nodes as a cue will be likely to include the other. So even if her attempted remembrance is accurate—here, <birthday>–<party>–<wine>–<restaurant>—it will be impossible to tell which experience of wine being served at a party was a contributing cause to the representational pattern.

Distributed network accounts of memory traces do not provide a way to track the causal history of memories for particular past events. This makes trouble for any attempt to incorporate such traces into the CTM, where excluding relearning is critical to the analysis of what remembering a particular past event requires. In cases of relearning, a person learns something, forgets it, and then learns it again. Representations of an event produced by remembering and representations of that event produced by relearning can only be distinguished by their causal history—when they were acquired relative to the event being represented. Distributed traces do not allow such distinctions to be made. They cannot be used to solve the challenge of relearning.

Both Bernecker and Michaelian acknowledge that an account of the CTM must exclude cases of relearning, and include conditions in their respective analyses to do so. As discussed in Sect. 5.2, they each claim that relearning is excluded because the causal connection between the past event and its subsequent representation must run through a memory trace. But they understand these memory traces to be distributed—that is, memory traces are superpositional and blended. The patterns that can be reconstructed from them reflect the influence of many past events, including possibly cases of relearning. Such traces cannot do the job to which they have been assigned. Bernecker and Michaelian must choose between their allegiance to the CTM and their commitment to distributed traces.

Those seeking to defend the compatibility between the CTM and distributed traces will obviously want to resist the conclusion that distributed traces and the CTM are incompatible. They may attempt to do so by claiming that a memory trace is stored not as a representation but as a *disposition* to form a representation, where "the disposition results from modifications to the strength of connections among the nodes in the relevant network."[14] Such a disposition would uniquely identify the past event, allowing for a distinction between remembering and relearning while maintaining the distributed network account of trace representations.

The possibility of dispositional traces is intriguing, but the success of such a proposal lies in the details—details that (to my knowledge) have not yet been made available. Is possessing a dispositional trace merely retention of the ability to represent the event in question? If so, then the account sounds like a reiteration of what Martin and Deutscher claimed, at least once their final condition is weakened to the Mental Representation Condition (as proposed in Sect. 4). There would then be no need to present a new analysis of remembering, nor endorse an account of traces as distributed patterns. Much as Martin and Deutscher did not need to include the additional commitment to structural analogues, contemporary CTM proponents could jettison the commitment to distributed traces. If the distributed patterns are considered essential to the updated CTM, then further details will be needed to understand how the disposition that serves as the trace is related to the pattern that is represented in the network. Is the disposition a set of instructions for how to recreate the event's pattern? If so, then it is hard to see how this disposition is not a representation. Is the disposition held as a set of connection strengths in the network—connections that would be made if any of the trace's nodes were activated? If so, then it seems that these dispositions will remain susceptible to the concerns above. The trace's particular dispositional pattern will fade, thereby shedding its unique causal signature.

Perhaps there are other ways for the incompatibility between the CTM and distributed traces to be avoided. There may be other conditions in Bernecker or Michaelian's analysis that can come to the rescue. For example, each of them builds in a condition that allows for subtle shifts in a memory's content over time. Could these small alterations provide the needed refinement? Such a strategy is unlikely to be successful. The addition of another condition will not remove the basic tension between the CTM's commitment to distinguishing traces by their causal history and the endorsement of an account of memory traces that does record such histories.

[14] I am grateful to an anonymous reviewer, and presumed defender of the CTM, for pressing this interpretation of the commitment.

Allowing shifts in content into the analysis of remembering may of course have other benefits. It may best reflect how our memory systems work, and so there may be good reason to endorse distributed traces that can account for this. Given the account's popularity with memory scientists, this view of memory traces may be the right one. Nothing I have said in this paper should be construed as an argument against that claim. The point I want to emphasize is that one cannot endorse this view of memory traces *and* the CTM simultaneously. The view of traces is popular with memory scientists precisely because its superpositional encoding makes individual memories hard to distinguish. Blended storage is thought to be the best explanation of the frequency of human misremembering errors. Here it is also worth noting that there are philosophers who endorse this distributed account of memory traces while abandoning the CTM—e.g., De Brigard (2014). The plausibility of such traces is not being questioned, only their compatibility with the CTM. If the distributed account of memory traces is the right one, then the CTM fails to provide an adequate analysis of memory.

7 The future of representing the past

The Causal Theory of Memory, as sketched by Martin and Deutscher (1966), provides a set of constraints on the connection between an initial event and its subsequent remembering. The account has intuitive pull. But the project is predicated on the possibility of tracing the unique causal influence of a particular past event up until the time that it is remembered. To achieve this, memory must be structured so as to retain discrete traces for each past experience a person is capable of recalling. However intuitive the account may be, it is in conflict with the view of memory's organization now favored by memory science. Memories are not stored discretely, it is claimed; instead, traces of individual events are distributed throughout a shared network of event features. The view's popularity is built around the assumption that such distributed traces provide the best way to explain the pervasiveness of memory errors. The trouble arises when contemporary memory theorists like Bernecker (2010) and Michaelian (2011) attempt to endorse the CTM *and* this distributed account of memory traces simultaneously. By appealing to an account of memory traces as distributed patterns, Bernecker and Michaelian are rejecting the view of memory's organization upon which the CTM relies. If memory does not keep track of particular past experiences, then there is no way to track the unique causal influence of such experiences as the analysis demands.

The basic conflict between the CTM and distributed memory traces has gone unnoticed because of the failure to distinguish between the various representational questions one can ask about memory traces. The Content Determination Question must be separated from the Memory Trace Question. Now that it is clear which question is being asked, we can go in search of an answer that—in one way or another—removes the tension between remembering's analysis and its evidence. In moving forward, both warrant further scrutiny. Does the CTM reflect our best analysis of what remembering requires? It is worth exploring whether the distinction from relearning is as central to our concept of remembering as has traditionally been supposed. It is also worth questioning whether distributed traces are in fact the best account of the process of remembering. Distributed traces may explain our tendency for memory errors, but we can ask whether they do equally well at explaining memory's successes, especially if successful remembering can occur for events that lack the recency and frequency that its networks privilege.

If such investigations reveal that either the analysis or the evidence are best understood differently, this would remove the apparent tension between remembering and memory traces. Or it may turn out that the incompatibility is real. The result would be disheartening, but informative nonetheless. We would know what our concept of remembering involves—and also that our minds are not designed for remembering.

Acknowledgments I am grateful to Sven Bernecker, Carl F. Craver, John Heil, Muhammad Ali Khalidi, and the audience at a University of Kansas colloquium for helpful comments on previous drafts of this paper. Special thanks to Corey J. Maley for creation of the figures.

References

Bennett, M., & Hacker, P. M. S. (2003). *Philosophical Foundations of Neuroscience*. Oxford: Basil Blackwell.

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

Bernecker, S., & Dretske, F. (2000). *Knowledge: Readings in contemporary epistemology*. Oxford: Oxford University Press.

Broad, C. D. (1925). *The mind and its place in nature*. London: Routledge and Kegan Paul.

De Brigard, F. (2014). Is memory for remembering? Recollection as a form of episodic hypothetical thinking. *Synthese*, *191*, 1–31.

Debus, D. (2010). Accounting for epistemic relevance: A new problem for the causal theory of memory. *American Philosophical Quarterly*, *47*, 17–29.

Deutscher, M. (1989). Remembering "Remembering.". In J. Heil (Ed.), *Cause, mind, and reality* (pp. 53–72). Dordrecht: Kluwer.

Grice, H.P. (1961). The causal theory of perception. *Proceedings of the Aristotelian Society*, 35 Supp., 121–152.

Heil, J. (1981). Does cognitive psychology rest on a mistake? *Mind*, *90*, 321–342.

James, W. (1890). *The Principles of Psychology*. London: Macmillan.

Laurence, J. R., & Perry, C. (1983). Hypnotically created memory among highly hypnotizable subjects. *Science*, *222*, 523–524.

Lewis, D. (1980). Veridical hallucination and prosthetic vision. *Australasian Journal of Philosophy*, *58*, 239–249.

Locke, D. (1971). *Memory*. London: Macmillan.

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

Mendelsohn, A., Chalamish, Y., Solomonovich, A., & Dudai, Y. (2008). Mesmerizing memories: Brain substrates of episodic memory suppression in posthypnotic amnesia. *Neuron*, *57*, 159–170.

Michaelian, K. (2011). Generative memory. *Philosophical Psychology*, *24*, 323–342.

Palmer, S. E. (1978). Fundamental aspects of cognitive representation. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and Categorization* (pp. 259–303). Hillsdale, NJ: Lawrence Erlbaum Associates.

Rosen, D. A. (1975). An argument for the logical notion of a memory trace. *Philosophy of Science*, *42*, 1–10.

Russell, B. (1912). *The Problems of Philosophy*. London: Williams and Norgate.

Shoemaker, S. (1970). Persons and their pasts. *American Philosophical Quarterly*, *7*, 269–285.

Smolensky, P. (1991). Connectionism and the language of thought. In B. Loewer & G. Rey (Eds.), *Meaning in Mind: Fodor and his Critics* (pp. 201–227). Oxford: Basil Blackwell.

Sutton, J. (1998). *Philosophy and Memory Traces: Descartes to Connectionism*. Cambridge: Cambridge University Press.

Von Eckardt, B. (1996). *What is Cognitive Science?*. Cambridge, MA: MIT Press.

Wixted, A., Squire, L. R., Yoonhee, J., Papesh, M. H., Goldiner, S. D., Kuhn, J. R., et al. (2014). Sparse and distributed coding of episodic memory in neurons of the human hippocampus. *Proceedings of the National Academy of Sciences*, *111*, 9621–9626.

Zemach, E. M. (1983). Memory: what it is, and what it cannot possibly be. *Philosophy and Phenomenological Research*, *44*, 31–44.