European Journal of Parapsychology
Volume 21.2, Special Issue, pages 182–202

ISSN: 0168-7263

Memory Without a Trace

Stephen E. Braude
Department of Philosophy,
University of Maryland, Baltimore County, USA

Abstract

Ever since Plato proposed that memories are analogous to impressions in wax, many have suggested that memories are formed through the creation of traces, representations of the things remembered. That is still the received view among most cognitive scientists, who believe the remaining challenge is simply to determine the precise physical nature of memory traces. However, there are compelling reasons for thinking that this standard view of memory is profoundly wrongheaded — in fact, disguised nonsense. This paper considers, firstly, what those reasons are in detail. Secondly and more briefly, it considers how trace-like constructs have undermined various areas of parapsychological theorizing, especially in connection with the evidence for postmortem survival-for example, speculations about cellular memory in transplant cases and genetic memory in reincarnation cases. Similar problems also emerge in areas often related to parapsychology — for example, Sheldrake's (1981) account of morphic resonance.

Introduction

One of the most persistent conceptual errors in philosophy, psychology, and neurophysiology is the attempt to explain memory by means of memory traces (sometimes called "engrams"). The underlying problems are very deep and difficult to dispel, and as a result, trace theories are quite seductive. In fact, in the cognitive sciences this approach

Correspondence details: Stephen E. Braude, Department of Philosophy, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, USA. Email: braude@umbc.edu.

to memory is ubiquitous and is almost never seriously questioned (for representative samples of the view, see, e.g., Damasio, 1996; Gazzaniga et al., 1998; Moscovitch, 2000; and Tulving and Craik, 2000). If doubts are raised at all, they typically concern how trace mechanisms are implemented or what the physical substrate of traces might be, not whether something is profoundly wrongheaded about the very idea of a memory trace. Moreover, positing memory traces is one aspect of a larger explanatory agenda that prevails in the behavioral sciences — namely, the tempting but ultimately fruitless strategy of explaining human behavior as if it is emitted by, and wholly analyzable in terms of, processes occurring within an agent. And one reason that agenda is so difficult to overturn is that in order to present a viable alternative, one must outline a very different approach to the analysis and understanding of human behavior.

But that last task goes well beyond the scope of this paper. My more modest goals here are (1) to summarize the main reasons for thinking that the concept of a memory trace is, not simply useless, but actually incoherent, and (2) to show, only briefly, how analogous concepts have crept insidiously into various areas of parapsychological theorizing, especially in connection with the evidence for postmortem survival — for example, speculations about cellular memory in transplant cases and genetic memory in reincarnation cases. Similar problems also undermine theorizing in areas often related to parapsychology — for example, Sheldrake's (1981) account of morphic resonance.

Why traces?

Suppose I meet my old friend Jones, whom I haven't seen in twenty-five years. How is it, we wonder, that I'm able to remember him? Many believe that I couldn't possibly remember Jones without there being something *in* me, a trace (presumably a modification in my brain), produced in me by my former association with Jones. Without that trace, that persisting structural modification in my brain, we'd apparently have causation over a temporal gap. We'd have to suppose that I remember Jones now simply because I used to know him. And to many, that looks like magic. How could something twenty-five years ago produce a memory now, unless that twenty-five-year gap is somehow bridged? So when I remember Jones after twenty-five years, we're tempted to think it's because something in me now closes that gap, linking my present memory to my past acquaintance with Jones.

Now parenthetically, I have to say that it's at least controversial (and in many instances rather naive) to suppose there's something wrong with the idea of causation over a temporal gap. Gappy causation is a problem only on the assumption that the only real causes are proximate causes (i.e., that cause and effect must be spatiotemporally contiguous). But that's a thread I can't pursue here. Positing memory traces is problematic enough quite apart from its underlying questionable picture of causation.

So, let's return to the motivation for asserting the existence of memory traces. Notice that traces aren't posited simply to explain how I happen to be in the particular states we identify as instances of remembering — for example, my experiencing a certain mental image of Jones. They're supposed to explain how memory is *possible* in the first place. The idea is that without a persisting structural modification in me, caused by something in my past — in this case, presumably, a physiological representation of Jones, no state in me *could* be a memory of Jones. So if after twenty-five years I have a mental image of Jones, the only way that image could count as a memory of Jones would be if it had the right sort of causal history. And the right sort of causal history, allegedly, is one that spatially and temporally links my present experience with my past acquaintance with Jones. So my image of Jones counts as a memory of Jones only if (1) there's a trace in me, caused by my previous acquaintance with Jones, and (2) the activation of that trace is involved in producing my present image of Jones. So mental images of Jones might be possible without that sort of causal history, but they wouldn't then be instances of remembering.

History has proven that this general picture of remembering is initially very attractive. But it gets very ugly very quickly, as soon as one asks the right sorts of questions. (In my view, this is where philosophy is most useful, and often the most fun: showing how claims which seem superficially plausible crumble as soon as their implications or presuppositions are exposed). What eventually becomes clear is that the idea of memory as involving *storage* is deeply mistaken, and that the mechanism of storage, memory traces conceived as representations of some kind, can't possibly do the job for which they're intended. This is actually an enormous topic and one of the most interesting subjects in the philosophy of mind. But since this issue is both vast and only part of what I want to discuss, I can't do more here than outline a few of the problems with the concept of a memory trace and indicate where one might look for additional details. (For extended critiques, see Bennett and Hacker, 2003; Braude, 2002; Bursen, 1978; Heil, 1978; Malcolm, 1977.)

More preliminaries

The first thing to note is that the problems with the concept of a memory trace are *hardware-independent*. It doesn't matter whether traces are conceived as mental or physical, or more specifically as static, dynamic, neurological, biochemical, atomic, subatomic, holographic (á la Pribram), nonspatial mental images, or (as Plato suggested) impressions in wax. No matter *what* memory traces are allegedly composed of or how they're purportedly configured, they turn out to be impossible objects. Memory trace theory requires them to perform functions that nothing can fulfill. So my objections to trace theory have nothing to do specifically with the fact that those theories are typically physiological or physical. Rather, it's because they're *mechanistic* and (in particular) because the mechanisms they posit can't possibly do what's required of them.

Before getting into details, I must deflect a certain standard reaction among scientists to the sort of criticisms I'm making here. Many have complained to me that, as scientists, they're merely doing empirical research, and so it's simply beside the point to argue, *a priori*, that their theories are unintelligible or otherwise conceptually flawed. However, I'm afraid that this response betrays a crucial naivete about scientific inquiry. There is no such thing as a purely empirical investigation. Every branch of science rests on numerous, often unrecognized, abstract (i.e., philosophical) presuppositions, both metaphysical and methodological. These concern, for example, the nature of observation, properties, or causation, the interpretation, viability, and scope of certain rules of inference, and the appropriate procedures for investigating a given domain of phenomena. But that means that the integrity of the discipline as a whole hinges on the acceptability of its root philosophical assumptions. If those assumptions are indefensible or incoherent, that particular scientific field has nothing to stand on, no matter how attractive it might be on the surface. And I would say that several areas of science, as a result, turn out simply to be bad philosophy dressed up in obscurantist technical jargon, so that the elementary nature of their mistakes remains well-hidden. Memory trace theory is just one example of this. And I'd argue that today's trace theories of memory, for all their surface sophistication, are at bottom as wrongheaded and simplistic as Plato's proposal that memories are analogous to impressions in wax. In short, I'd say they are disguised nonsense.

Two more disclaimers, before outlining my objects to trace theory. First, when I say that the concept of a memory trace is incoherent or that trace theory is conceptually naive in certain respects, I'm not saying that trace theories — or the scientists who hold them — are stupid. To say that a proposal or concept is nonsensical or incoherent is simply to say it makes no sense. Now although the world isn't suffering a shortage of stupidity, not all nonsense is stupid. In fact, the most interesting nonsense is *deep* nonsense, and it's something which can all too easily deceive even very smart people. That's because the problematic assumptions are buried well below the surface and require major excavation.

Second, I've learned over the years that when I outline my objections to trace theory, many hear me as suggesting that the brain has nothing to do with memory. I'll say a bit more about this later, but for now I'll just note that I'm saying nothing of the kind — although evidence for postmortem survival *would* seriously challenge this. In fact, let's overlook for now complications to all physiological cognitive theories posed by the evidence for postmortem survival and restrict our attention to embodied humans. In those cases, clearly, the capacity to remember is causally dependent, not simply on having a functioning brain, but probably also on changes to specific areas of the brain. However, it's one thing to say that the brain *mediates* the capacity to remember, and another to say it *stores* memories. The former view (more likely the correct one) takes the brain to be an instrument involved in the expression of memory; the latter view turns out to be deeply unintelligible. For a very limited analogy, we can say that while a functionally intact instrument may be causally necessary for performing a musical improvisation, the music is not stored in the instrument (or anywhere else).

The horns of a dilemma

So why is the concept of a memory trace incoherent? Let's begin with an analogy (drawn from John Heil's outstanding critique of trace theory — Heil, 1978). Suppose I invite many guests to a party, and suppose I want to remember all the people who attended. Accordingly, I ask each guest to leave behind something (a trace) by which I can remember them. Let's suppose each guest leaves behind a tennis ball. Now clearly I can't use the balls to accomplish the task of remembering my party guests. For my strategy to work, the guests must deposit something reliably and specifically linked to them, and the balls obviously aren't differentiated and unambiguous enough to establish a link only with the person who left it.

So perhaps it would help if each guest signed his/her own tennis ball, or perhaps left a photo of him/herself stuck to the ball. Unfortunately, this threatens an endless regress of strategies for remembering who attended my party. Nothing reliably (much less uniquely and unambiguously) links the signature or photo with the guest who attended. A guest could mischievously have signed someone else's name, or left behind a photo of another person. Or maybe the signature was illegible (most are), or perhaps the only photo available was of the person 25 years earlier (e.g., when he still had hair, or when he had a beard, wore eyeglasses, and was photographed outdoors, out of focus and in a thick fog), or when he was dressed in a Halloween costume or some other disguise.

But now it looks like I need to remember in order to remember. A tennis ball isn't specific enough to establish the required link to the person who left it. It's not the sort of unambiguous representational calling card the situation requires. So we supposed that something else might make the tennis ball a more specific link — a signature or a photo. That is, we tried to employ a secondary memory mechanism (trace) so that I could remember what the original trace (the tennis ball) was a trace of. But the signature and photo are equally inadequate. They too can't be linked unambiguously to a specific individual. Of course, if I could simply *remember* who wrote the signature or left behind the photo, then it's not clear why I even needed the original tennis balls. If no memory mechanism is needed to make the connection from tennis ball to party guest, or illegible signature to its author, then we've conceded that remembering can occur without corresponding traces, and then no trace was needed in the first place to explain how I remember who attended my party. So in order to avoid that fatal concession, it looks like yet another memory mechanism will be required for me to remember who left behind (say) the illegible or phony signature, or the fuzzy photo. And off we go on a regress of memory processes. It seems that no matter what my party guests leave behind, nothing can be linked only to the guest who left it. We'll always need something else, some other mechanism, for making the connection between the thing left behind and the individual who left it.

In fact, it seems that the only way to stop the regress is for a guest to leave behind something that is *intrinsically* and exclusively linked only to one individual. That is why Wolfgang Köhler, for example, proposed that traces must be *isomorphic* with the things of which they're traces — that is, the things they represent (e.g., Köhler, 1947, 1969). But what Köhler and others have failed to grasp is that this kind of intrinsic connection is impossible, because nothing can function in one and only one way. As I'll argue shortly, this is especially clear when the function in question is one of representation or meaning. Nothing can represent unambiguously (or represent one and only one thing); representing is not something objects can do all by themselves; and representation can't be an intrinsic or inherent relation between the thing represented and the thing that represents it.

Interestingly, although Köhler failed to see why trace theory is doomed to fail, he was remarkably clear about what trace theory requires. Köhler understood that a major hurdle for trace theory is to explain trace *activation* — that is, how something present triggers my trace of Jones, rather than the trace of someone else. And that's a serious problem, because what triggers a memory (or activates a trace) can be quite different from what established it in the first place. So Köhler wrote,

> "...recognition... means that a present fact, usually a perceptual one, makes contact with a corresponding one in memory, a trace, a contact which gives the present perception the character of being known or familiar. But memory contains a tremendous number of traces, all of them representations of previous experiences which must have been established by the processes accompanying such earlier experiences. Now, why does the present perceptual experience make contact with the *right* earlier experience? This is an astonishing achievement. Nobody seems to doubt that the *selection* is brought about by the similarity of the present experience and the experience of the corresponding earlier fact. But since this earlier experience

> is not present at the time, *we have to assume that the trace of the earlier experience resembles the present experience, and that it is the similarity of our present experience (or the corresponding cortical process) and that trace which makes the selection possible*."
>
> (Köhler, 1969, p. 122, emphasis added)

By the way, this passage reveals another serious limitation of trace theory, one I can only mention in passing here. If trace theory has any plausibility at all, it seems appropriate only for those situations where remembering concerns past *experiences*, something which apparently could be represented and which also could resemble certain triggering objects or events later on. But we remember many things that aren't experiences at all, and some things that aren't even past — for example, the day and month of my birth, the time of a forthcoming appointment, that the whale is a mammal, the sum of a triangle's interior angles, the meaning of "anomalous monism." Apparently, then, Köhler's point about trace activation and the need for similarity between trace, earlier event, and triggering event, won't apply to these cases at all. So even if trace theory was intelligible, it wouldn't be a theory about memory generally.

In any case, trace theory is not intelligible, and Köhler's observation reveals why. To avoid the circularity (and potential regress) of positing the ability to remember in order to explain my ability to remember (e.g., by requiring further trace mechanisms to enable the previous trace do its job), we must suppose that some trace uniquely and unambiguously represents or connects to the original experience. And because unambiguous representation is an impossible process, trace theory is caught between two fatal options. I'll explain in a moment why unambiguous representation is impossible, but first, we need to observe that the tennis ball/party example hides a further complication noted in the passage from Köhler.

Traces are usually supposed to be brain processes of some sort, some physiological representation produced, in this case, by a party guest. But what *activates* this trace later can be any number of things, none of which need to resemble the experience, object, or event that produced the original trace. Suppose Jones attended my party. Trace theory requires my experience of Jones at the party to produce a representation in me of Jones (or my experience of him) so that I can later remember that he was at the party. But what will subsequently activate that trace?

It could be Jones himself, or an image of Jones, or the lingering smell of someone's cologne, or a telltale stain on the carpet, or perhaps someone asking, "Who was at the party?" Of course, some of these potential triggering objects or events might plausibly be said to resemble the thing that originally produced the trace. But how can (say) the smell of cologne, a stain, or the words "Who was at the party?" trigger the trace of Jones created by his presence at the party? These things aren't obviously similar to Jones himself. If we posit another memory *mechanism* to explain how I draw the connection between the cologne and Jones (e.g., he may have worn it, spilled it, or simply talked about it), or how the question "Who was at the party?" leads me to the right party and not some other party, or even how I remember what the word "party" means, we're starting a regress of memory mechanisms. But if we say it's because I can simply *remember* who wore (or perhaps mentioned) the cologne, stained the carpet, or who my party guests were, then we're still reasoning in a circle. We're still explaining memory by appealing to the ability to remember. Moreover, if I can remember these things without some further trace, then we didn't need a trace in the first place to explain my ability to remember that Jones was at the party. However, if we follow Köhler's lead, then we have to assert some kind of intrinsic similarity or resemblance, some kind of psychophysical *structural isomorphism*, between three things: the original experience or event, the trace produced on that occasion, and the subsequent triggering events.

If nothing else, it should make you suspicious that a representation of Jones at the party will be isomorphic both to Jones (or my experience of him) and to the innumerably many and quite different things that can later activate the trace — for example, a particular scent or a sequence of sounds. What kind of similarity could this be? The answer is that it can't be any kind of similarity and that Köhler's proposal is literally meaningless. As tempting as it is to continue for a while enumerating the problems with trace theories, I'll restrict myself now to two more points, to explain perhaps the deepest confusion underlying these theories.

The first problem is with the very idea of structural isomorphism. The term "structural isomorphism" sounds impressive and scholarly, but in trace theories the appeal to structural isomorphism is really just the appeal to an *inherent similarity* between two things, *determined solely by their respective structures*. Traces must be produced in a way that relates them structurally to the things of which they're traces, and they must be activated only by things having the right underlying structure. Moreover, that activation must be determined solely by intrinsic relations between the structures of the trace and the things that activate them. Otherwise, we'd need another mechanism to explain how the *right* trace is activated in the presence of a trigger that could just as well have been isomorphic with (or mapped onto) something else. And that raises the circularity or regress problem noted earlier.

But the alternative, inherent similarity, makes no more sense than saying that a square is a circle. Inherent similarity is a *static* relation obtaining only between the similar things. And it must hold between those things *no matter what*. If, for example, context could alter whether two things count as similar, then those things are not similar merely in virtue of intrinsic relations holding between their respective structures. But that's why intrinsic similarity is nonsense. Similarity exists only with respect to variable and shifting criteria of relevance. It can only be a dynamic relation holding between things at a time and within a context of needs and interests.

A simple example from geometry should make the point clearly. Consider the five geometric figures in Figure 1.

Now consider the question: To which of the last four figures is the triangle (a) similar? The proper response to that question should be puzzlement; you shouldn't know how to answer it. Without further background information, without knowing what matters in our comparison of the figures, the question has no answer at all. Mathematicians recognize this, although instead of the term "similarity" they use the expression "congruence." In any case, mathematicians know that in the absence of some specified or agreed-upon rule of projection, or function for mapping geometric figures onto other things, no figure is congruent with (similar to) anything else.

Mathematicians recognize that there are different standards of congruence, appropriate for different situations. But no situation is *intrinsically basic*, and so no standard of congruence is inherently privileged or more fundamental than others. For example, engineers might sometimes want to adopt a fairly strict mapping function according to which (a) is congruent only with other figures having the same interior angles and the same horizontal orientation. But in that case, (a) would be congruent with none of the other four figures. Of course, only in very specialized contexts are we likely to compare figures with respect to their horizontal orientation. In many situations it would be appro-

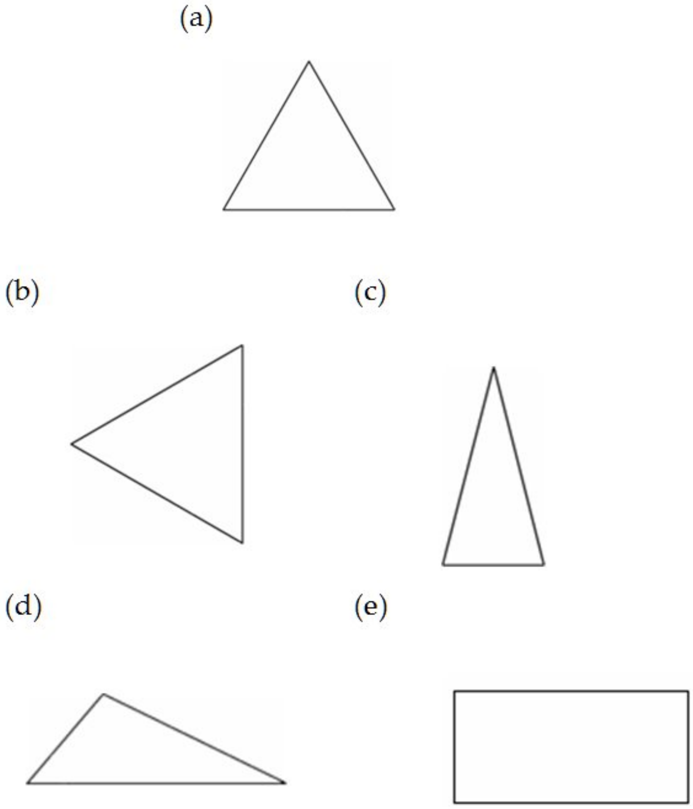

Figure 1. Five simple geometric figures

priate to adopt a different standard of congruence, according to which sameness of interior angles is all that matters. And in that case we'd say that figures (a) and (b) are congruent but that (a) is not congruent with the other figures. However, there's also nothing privileged about sameness of interior angles. Perhaps what matters is simply that (a) is congruent with any other three-sided enclosed figure, in which case we could say it's congruent with the three triangles (b)–(d), but not with the rectangle (e). But even that criterion of congruence can be modified or supplanted. Mathematicians have rules of projection that map triangles onto any other geometric object, but not to (say) apples or oranges. Of course, the moral here is obvious. If simple geometric figures are not intrinsically similar — that is, if they count as similar only against a background of assumptions about which of their features matter (i.e., are relevant), then we certainly won't find intrinsic similarity with much more complex objects — in particular, memory traces and the various objects or events that allegedly produce and activate them.

But maybe you're still not convinced. Perhaps you think that there *is* a fundamental principle of congruence for this geometric example. You might think that, first and foremost, (a) is similar to just those figures with sides of exactly the same length, the same horizontal orientation, and with exactly the same interior angles. And perhaps you'd want to call that something like "strict congruence (or identity)." But there are at least three serious problems with that position.

First, even if this sort of congruence counted as more fundamental than other forms of geometric similarity, that could only be in virtue of a kind of historical accident. The primacy of that standard of congruence would reveal more about us, our conventions and values — in short, what merely happens to be important to us, than it does about the figures themselves. In fact, it's a standard appropriate for only a very narrow range of contexts in which we consider whether things are similar. Second (and as an illustration of that first point), it's easy to imagine contexts in which two triangles have exactly the same interior angles, horizontal orientation, and sides, but don't count as similar. If we're interior designers, for example, it might also matter whether the triangles are of the same color, or whether they're placed against the same colored background, or whether they're made of the same material. If we're graphic artists, it might matter whether the triangles were both original artworks or whether one was a print. Or if we're librarians or archivists, it might matter whether the triangles occur on the same page of different copies of the same book. And third, even if we could decide on some very strict sense of congruence (or identity) which would count as privileged over all other forms of similarity, it would be useless in the present context. Memory traces are never strictly identical either with the things that produce them or with the things that activate them. The looser and more complex forms of similarity at issue in trace theories are classic examples of the sorts of similarities that can't possibly be inherent, static relations between things.

And as if that weren't enough, another aspect of this general confusion about similarity is the requirement that traces and other things have intrinsic or inherent *structures* — that is, some context-independent parsing into basic elements. Because isomorphism (mapping) is tied to structural elements of the isomorphic things, that's a necessary condition for intrinsic isomorphism to hold between the trace and the things it represents. After all, if what counted as structure depended on context — that is, if a trace could just as well have been parsed differently and assigned alternative structures, then it could be mapped onto (or count as similar to) different things. And unfortunately for trace theory, objects and events can always be parsed in an indefinite number of ways, and whatever parsing we select can only be conditionally, and never categorically or intrinsically, appropriate. We always determine a thing's components relative to a background against which certain features of the things (but not others) count as relevant. But then it's only against shifting and non-privileged background criteria of relevance that we take two things to have the same structure; they are never isomorphic *simpliciter*.

So the trace theorist's inevitable appeal to privileged, inherent structures and intrinsic mappings is literally absurd. It's on a par with claiming that a pie has a basic context-independent division into slices or elements, or that there's an absolutely context-independently correct and privileged answer to the questions, "How many events were there in World War II?" and "How many things are in this room?"

Confusions about representation

The appeal to inherent similarity or structure is merely a specific form of a more pervasive problem in the so-called cognitive sciences — namely, confusions about and equivocations on the term "representation." Traces are supposed to represent their causes, the events or experiences that produced them, and they must be internally and structurally differentiated in ways that correspond to the different things we remember. This is one version of the general view that distinct mental states are caused by (or are identical to) certain corresponding distinct internal physical states, and that what those different internal states *are* (i.e., what they represent) is explainable wholly in terms of their distinctive structural features. At this point, cognitive scientists typically do a lot of hand waving and say something like, "We may not currently know all the details, but presumably some super psychology of the future (or perhaps God) could in principle look inside our heads and know, from the way we're configured, what we're thinking."

However, this general picture rests on the utterly false assumption that a thing's representational properties can be determined solely by its structural or topological features. I've examined this error in considerable detail elsewhere (Braude, 1997, 2002). For now, a few brief remarks will have to suffice.

To see what's wrong, we need to appreciate that *anything can represent anything*. In fact, a thing's representational options are limited only by the situations into which it can be inserted. And if that's the case, then what something represents can't simply be a function of how it's configured. Things must be *made* to represent or mean something. Suppose I'm trying to teach a child the alphabet. I show him a picture of a dog and I say "D is for dog." In that case, we might say that the picture represents the class of dogs. But I could have said, "C is for collie," and in that case the picture would have represented a subset of the set of dogs. Similarly, I could have said "L is for Lassie," in which case the picture would have represented an even smaller subset of dogs. I could also have said "Z is for Ziggie," referring to the child's pet collie. And notice, these changes in what the picture represents have nothing whatever to do with corresponding changes in the arrangement of pixels, or atoms, or anything else in the picture. Those structural features of the pictures remained the same in all cases. What the picture represents depended instead on how it was used.

And in fact, the picture's representational properties could be changed even more dramatically. My disgruntled students could make the picture represent me and symbolically express their hostility toward me by using it as target for darts. Or I could jokingly point to the picture and say "This was Joan Rivers before plastic surgery." Or suppose I'm trying to give directions to someone without the aid of a map. I could place the picture on a table and say, "This is the shopping center, this [a ham sandwich] is the hospital, this [my fork] is the access road, and this [a salt shaker] is the water tower."

Of course, contexts in which (say) a sandwich represents a building, or in which a picture of a dog represents a distinguished philosopher (or over-the-hill comedienne), are atypical in some respects. But those situations are unusual *only* with respect to what the objects represent. They aren't at all unusual with respect to how representational properties are acquired. And it doesn't matter whether we're talking about images, words, or (say) synaptic connections. In every case (familiar and offbeat), what a thing represents depends ultimately on the way we place it in a situation. There are no context-independent forms of representation or meaning. So when it comes to examples like the picture of a dog or the ham sandwich, the mistake many make is to think that some representational properties — the familiar and apparently default ones — are inherently fundamental and that others are anomalous. That is, they believe that representation in familiar cases is somehow built-into or hardwired into the representing objects, and that this inherent function simply gets *overridden* in the more unusual cases. But in fact, the familiarity of certain contexts reveals more about us, about our patterns of life and our interests, than it does about the objects themselves. If our form of life were radically different, the default or familiar representational properties of objects could change accordingly.

But then if a brain structure (say) is to represent something past and function as a memory trace, it can't do so solely in virtue of its structural features. Nothing represents or means what it does on topological grounds alone. However, the whole point of Köhler's principle of psychophysical isomorphism (or related hypotheses in the cognitive sciences) is to tie what a thing represents solely to its structure. That was the only way to avoid the equally fatal error of requiring a regress of mechanisms to explain how the original mechanism or state can do its job. So this, too, turns out to be a dead end.

Tokens and types

But let's return more explicitly to trace theory. A related, and equally unheralded problem with such theories is that traces and their causes or activators are of radically different ontological kinds, and the sort of thing traces have to be is a kind that many think is simply a philosophical fiction. At any rate, it's nothing but a philosophical move, not even remotely a scientific move, to posit the existence of traces. Hopefully, one distinction and one more example will make this clear.[1]

Trace theorists have always been tempted to regard traces as kinds of *recordings* of the things that produced them. In fact, some previous influential writings on memory compared traces to tape recordings or grooves and bumps in a phonograph record. The justification for that idea, as we've seen, is that traces must somehow capture essential structural features of the things that produce them. However, the poverty of this view is easy to expose.

Consider: One of the things I remember is Beethoven's Fifth Symphony (hereafter abbreviated as B5). Modern versions of trace theory require that my memory is explained in terms of a representation of B5, stored in some form in my brain and produced in me by the experience of hearing B5 in the past. This trace must have certain structural

[1] For a considerably more detailed presentation of the following arguments, see Bursen, 1978.

or topological properties that link it to the thing(s) that caused it, properties which also distinguish it from traces of other pieces of music. So presumably this trace of B5 was produced by and captures features of a performance I heard of B5. But which features? Tempo, rhythm, pitch, length of notes, instrumental timbre, dynamic shadings? You'd think so if my trace of B5 was produced by and represents or records a B5 performance, and also if that trace is to differ (say) from my trace of Beethoven's Fourth (B4) or "Yankee Doodle.". But I (like many others) can remember B5 by recognizing a wide variety of musical performances as *instances* (or as philosophers would put it, *tokens*) of B5. For instance, I could recognize B5 when certain notes are held for an unusually long time, or when it's played with elaborate embellishments, or with poor pitch and many mistakes by an amateur orchestra. In fact, I could recognize truly outlandish musical events as instances of B5 — for example, when it's played extremely slowly or rapidly, or with tempo changing every bar, or with arbitrary notes raised a major sixth, or when it's played with inverted dynamics, or played only on kazoos, banjos, or tubas. Similarly, I could recognize a series of percussive taps as a pitch-invariant version of the opening bars of B5.

But this means that the trace is not a recording. On the contrary, it must be a very unusual sort of entity. Whereas the remembered and triggering events or experiences are concrete event-tokens, the trace itself must be a relentlessly abstract object — what philosophers call a *type*. And it has to be so abstract that it can't contain *any* features found in the performances or experiences that produced it (e.g., precise rhythm, pitch, etc.). If it had those features, we'd need to posit another mechanism to explain how my trace can be activated by tokens of B5 lacking them — for example, a tuba-only performance of B5 played at quarter speed with many wrong notes. But if we try to prevent a regress by saying that I can simply recognize that the tuba-only version is an instance of B5, then we don't need to posit the trace of B5 at all. We've conceded that I can remember B5 without recourse to a B5 trace.

But now look at what has happened. We've seen that the B5 trace is an abstract type. However, that trace has to have *some* features in virtue of which it's a B5 trace and not a trace of (say) Beethoven's Fourth, the "Waldstein" Sonata, or "Yankee Doodle." But it can't have features found in any specific instances of B5, because none of those are necessary for a musical event to be an instance of B5 capable of either producing or activating the trace. So the B5 trace somehow needs to have features necessary and sufficient for being an abstract B5 and not (say) a B4, but without having any specific features regarding pitch, tempo, dynamics, etc., all of which can be changed or absent (perhaps you can now see why many consider abstract types impossible objects).

In any case, we've arrived at the point where we see the ultimately nonscientific nature of trace theory. It's committed to the view that a memory trace and all its concrete instances have a structure that is essential to all things that are instances of B5, but none of the specific features which versions of B5, including the nightmare versions, can lack. This position is commonly called *Platonic essentialism* — the view that things are of the same kind in virtue of sharing a common underlying, but abstract, structure. And that's not a scientific view at all. It's a philosophical view, and a bad one at that.

The abuse of memory in parapsychology

It's unfortunate enough that memory trace theory is received dogma in the cognitive sciences. Almost no one seems to doubt that memories are somehow stored and encoded in us. So it's not surprising that this picture of memory has found its way to more overtly speculative or frontier areas of science, including parapsychology. No doubt it's very tempting for parapsychologists to posit trace-like processes in their own theories, because they will at least appear to be reasoning along scientifically orthodox lines, even if the subject matter itself falls outside the scientific mainstream.

For example, Roll has proposed a "psi structure" theory of survival, modeled explicitly after memory trace theory, and according to which memory traces are left, not simply in individual brains, but in our environment as well (Roll, 1983). Of course, this escapes none of the classic problems of trace theory, because on Roll's view, what certain structures represent (or are similar to) remains unintelligibly tied to inherent features of those structures. This is especially problematical when Roll suggests that an individual mind or personality is a system of such structures. That's no more plausible than saying that we can tell whether a person is thinking about his grandmother just by examining the state of his brain, or that a picture of a dog represents something specific independent of its use in a context. It requires brain or mental structures to mean or represent something simply in virtue of how they're configured, never mind their dynamic position within an equally dynamic life situation. Roll also proposes explaining ESP as the responding to memory traces left on objects by previous guesses. But that seems no more credible than supposing that I could remember my party guests from looking simply at the tennis balls they left behind, or the illegible signatures or photos they left along with the balls.

Trace theory also appears in other guises in connection with the evidence for postmortem survival. One is the suggestion that reincarnation cases can be explained in terms of genetic memory. However, I've found no serious researcher making that suggestion. It seems, instead, to be entertained simply as a real possibility, albeit one that can be rejected on empirical grounds (see, e.g., Almeder, 1992; Stevenson, 1974). That is, it's treated as if it's an intelligible position that happens merely to be inadequate to the data. Another application of trace theory to survival is the attempt to explain transplant cases by appealing to cellular memory (e.g., Pearsall et al., 1999). No doubt the reason it's tempting here to posit genetic or cellular memory traces is that in reincarnation and transplant cases, complex psychological regularities seem to persist in the absence of the usual presumed bodily correlates. So to those for whom it's unthinkable that memories could persist without being stored somewhere, it might seem reasonable to propose that memories and personality traits can be encoded in a kind of hardware that has nothing to do with the brain. However, since the problems noted earlier with trace theories are hardware-independent, it's an insignificant change merely to relocate the traces in different physical systems. It's still untenable to suppose that representation, meaning, or similarity, are determined solely by a thing's topological features.

To me, it's interesting that when the usual suspect — the brain — isn't available as the locus of memory storage, some find it inevitable that memories must simply be located in a different place or perhaps in a modified form. It demonstrates just how deeply mechanistic assumptions have taken root, and in a way, it shows a profound lack of scientific imagination. The situation here closely parallels what happened in response to Lashley's famous experiments in the 1920s (Beach et al., 1960; Lashley, 1929, 1950). When Lashley found that no matter how much of a rat's brain he surgically removed, trained rats continued to run their maze, some concluded that the rats' memories weren't specifically localized in their brains. Instead, they suggested that the memories were diffusely localized, much as information is diffusely distributed in holograms (Pribram, 1971; Pribram et al., 1974; Pribram, 1977). But to someone not antecedently committed to traditional mechanistic dogma, Lashley's experiments take on a different sort of significance, perhaps similar to that of the evidence for postmortem survival. They suggest that memories are not located anywhere or in any form in the brain. And more generally, they suggest that the container metaphor (that memories and mental states in general are *in* the brain, or in something else) was wrong from the start. Of course, that's what my arguments in the preceding sections were intended to show.

Another variant of this general error emerges in Rupert Sheldrake's (1981) suggestion that morphic fields capture the essential structure of developmental forms and even behavioral kinds. Although Sheldrake thought he was escaping the evils of mechanistic theories with his view, in fact he retained the underlying errors of supposing that similarity is an intrinsic structural relation between things, and that things of the same kind are of that kind because they share a common underlying structural essence. The claim that behavioral kinds, such as feeding behavior and courtship, can be captured in strictly structural terms, is especially implausible.[2]

Summing up

I realize that I'm pretty much a voice in the wilderness on these issues, and I find myself in the unenviable position of having to argue that many prominent and respected scientists actually don't know what they're talking about. I wish there were some other, less fundamentally upsetting, way to undercut trace theories of memory. But I believe that the problems really are that deep and that the theories really are that essentially confused.

However, as long as I'm being antagonistic, I see no compelling reason to stop where I left off. I might as well finish with brief obnoxious coda. As I see it, both memory researchers and parapsychologists are missing an opportunity to be genuine scientific pioneers. Rather than boldly searching for new explanatory strategies (for memory specifically and for human behavior generally), they cling instead to familiar mechanistic presuppositions, which they've typically never examined in any depth, but by means of which they can maintain the illusion that they're doing science according to the allegedly tough-minded methods exemplified in some physical sciences. (Sherry Turkle has appropriately

[2] For a detailed critique of Sheldrake's theory, see Braude (1983).

called this “physics envy.”) They can’t get past the assumption that human abilities and behavior must be analyzed in terms of lower-level processes and mechanisms. And they seem not to recognize the difference between claiming that cognitive functions are *analyzable* in terms of underlying physical processes and claiming instead that those functions are merely *mediated* by underlying physical processes. But there are novel explanatory options and strategies they never consider; there are alternative and profoundly different approaches to the understanding of human beings. However, spelling out those options is a huge project, one that must be reserved for another occasion.

References

Almeder, R. (1992). *Death and Personal Survival*. Lanham, MD: Rowman & Littlefield.

Beach, F.A., Hebb, D.O., Morgan, C.T., and Nissen, H.W. (eds.) (1960). *The Neuropsychology of Lashley: Selected Papers of K. S. Lashley*. New York: McGraw-Hill.

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

Braude, S.E. (1983). ”Radical Provincialism in the Life Sciences: A Review of Rupert Sheldrake’s *A New Science of Life*.” *Journal of the American Society for Psychical Research*, 77, 63–78.

Braude, S.E. (1997). *The Limits of Influence: Psychokinesis and the Philosophy of Science*. Lanham, MD: University Press of America.

Braude, S.E. (2002). *ESP and Psychokinesis: A Philosophical Examination (Revised Edition)*. Parkland, FL: Brown Walker Press.

Bursen, H.A. (1978). *Dismantling the Memory Machine*. Dordrecht, Boston, London: D. Reidel.

Damasio, A.R. (1996). *Descartes’ Error: Emotion, Reason and the Human Brain*.

Gazzaniga, M.S., Mangun, G.R., and Ivry, R.B. (1998). *Cognitive Neuroscience: The Biology of the Mind*. New York: Norton.

Heil, J. (1978). ”Traces of Things Past.” *Philosophy of Science, 45*, 60–67.

Köhler, W. (1947). *Gestalt Psychology*. New York: Liveright.

Köhler, W. (1969). *The Task of Gestalt Psychology*. Princeton: Princeton University Press.

Lashley, K.S. (1929). *Brain Mechanisms and Intelligence*. Chicago: University of Chicago Press.

Lashley, K.S. (1950). ”In Search of the Engram.” *Symposia of the Society for Experimental Biology, 4*, 454–482.

Malcolm, N. (1977). *Memory and Mind*. Ithaca: Cornell University Press.

Moscovitch, M. (2000). ”Theories of Memory and Consciousness.” In E. Tulving and F.I.M. Craik (eds), *The Oxford Handbook of Memory* (p. 609-626). Oxford: Oxford University Press.

Pearsall, P., Schwartz, G.E.R., and Russek, L.G.S. (1999). ”Changes in Heart Transplant Recipients That Parallel the Personalities of Their Donors.” *Integrative Medicine,* 2, (2/3), 65–72.

Pribram, K.H. (1971). *Languages of the Brain* . Englewood Cliffs, N.J.: Prentice Hall.

Pribram, K.H. (1977). ”Holonomy and Structure in the Organization of Perception.” In U.M. Nicholas (Ed.), *Images, Perception and Knowledge*. Dordrecht: Reidel.

Pribram, K.H., Nuwer, M., and Baron, R.U. (1974). ”The Holographic Hypothesis of Memory Structure in Brain Function and Perception.” In D.H. Krantz, R.C. Luce, and P. Suppes (eds), *Contemporary Developments in Mathematical Psychology, Vol. 2*. San Francisco: Freeman.

Roll, W.G. (1983). ”The Psi Structure Theory of Survival.” In W.G. Roll, J. Beloff, and R. White (eds), *Research in Parapsychology 1982* (p. 117-120). Metuchen, NJ & London: Scarecrow Press.

Sheldrake, R. (1981). *A New Science of Life: The Hypothesis of Formative Causation*. London: Blond & Briggs Limited.

Stevenson, I. (1974). *Twenty Cases Suggestive of Reincarnation, 2nd Ed. Rev*. Charlottesville: University Press of Virginia.

Tulving, E. and Craik, F.I.M. (eds.) (2000). *The Oxford Handbook of Memory*. Oxford: Oxford University Press.