Personal identity and mental time travel

Marya Schechtman

Abstract
This paper examines the role of episodic memory, and the broader notion of “mental time travel” (MTT), in constituting personal identity. After arguing that the construal of memory’s role in personal identity found in traditional psychological continuity theories of personal identity is both unrealistic and unsatisfying, the paper endeavors to provide a better account. This begins with recent work in the science and philosophy of memory that sees episodic memory as part of a broader faculty for MTT (which also involves imagination and counterfactual thought). Some of the basic ideas expressed in this work are developed into an account of the connection between MTT with “strong identification” and personal identity. According to this alternative approach, we regularly “borrow affect” from our pasts and futures through forms of remembering, imagining, and counterfactual thinking that involve a particular form of identification with our past and future selves. This activity generates a strong diachronic experience of self, which contributes in important ways to diachronic personal identity. The sense of self generated through MTT is, however, only one piece of a more comprehensive account of personal identity. The paper concludes by describing its place in the larger picture.

Keywords
Imagination ∙ Memory ∙ Mental time travel ∙ Personal identity ∙ Self-experience

This article is part of a special issue on “Successful and Unsuccessful Remembering and Imagining”, edited by Ying-Tung Lin, Chris McCarroll, Kourken Michaelian, and Mike Stuart.

It is something of a truism that our memories “make us who we are” this thought is found in philosophy, literature, psychology, popular song lyrics, greeting cards, and in the Pixar film Inside Out. Its ubiquity suggests that it expresses something compelling, but it is not entirely obvious exactly what that is. Which memories make us who we are? Is there a special class, or is it all memories? What form of identity is at issue? Numerical identity, psychological identity, moral identity, social identity? And what is meant by the claim that our memories make us who we are as opposed, for instance, to the experiences remembered making us who we are while memories provide self-knowledge? I suspect that the claim that
memories make us who we are in fact means different things as asserted by different individuals in different contexts. In this paper I focus on the idea that memory is (at least partially) constitutive of personal identity as it is worked out in psychological continuity theories of personal identity such as those developed by Lewis (1976, p. 158), Parfit (1986), Perry (1976), and Shoemaker (1963). I argue that recent empirical and philosophical developments in the study of memory complicate the picture of memory presupposed in these theories, and hence their understanding of the connection between memory and personal identity, as well as suggesting a different way of thinking about this connection.

I begin by reviewing familiar features of psychological continuity theories to articulate the connection between memory and personal identity they presuppose. Next, I describe some developments in recent empirical and philosophical work on memory. I use these developments as a springboard, expanding some of the key ideas they contain to propose an alternative picture of the connection between memory (among other, related, cognitive functions) and personal identity. The path via which I will explore this alternative picture is somewhat winding, involving, among other things, an extended discussion of imagination. Rather than providing a more detailed overview here, I will signpost along the way.

1 Locke and psychological continuity

The idea that identity is constituted by memory is often associated with John Locke (1975). Locke does not actually say this; he says that identity is constituted by sameness of consciousness, telling us for instance, that “[…] as far as this consciousness can be extended backwards to any past Action or Thought, so far reaches the Identity of that Person” (1975, p. 335). He does seem to see memory as playing an important role, however. An early critic, Thomas Reid (1975, p. 115) suggests that there is nothing it could mean to say that someone has present consciousness of a past experience other than that he remembers it. One standard reading of Locke’s account is thus that he defines identity in terms of memory “from the inside”, or what has since come to be called “episodic” memory. Roughly, the idea is that the memories relevant to constituting personal identity come to us with imagery, affect, and a feeling of “reliving” a past experience rather than as propositional memories of facts about the past.

Psychological continuity theorists see Locke as offering important insights. Two, in particular, play an important role in their own views: first, the idea that personal identity should be defined in terms of psychological relations rather than sameness of substance, and second that facts about personal identity are inherently bound up with practical judgments (e.g., the assignment of praise or blame, and the rationality of egoistic concern). Locke describes this broad practical significance of identity by saying that “Person” is a “forensic” term (Locke, 1975, p. 346). Psychological continuity theorists agree, however, with the prevailing view that a simple memory theory, according to which whatever experience a person repre-
sents herself as having had in the past is, for that reason, truly her experience, is untenable. They therefore alter and update the memory theory to make it stronger.

One alteration is especially important for our purposes. Psychological theorists address a longstanding objection that a memory criterion of personal identity is circular. The circularity objection is familiar: A memory theory is only remotely plausible if we have a way to separate genuine memory from delusion or confabulation. If, I sincerely claim to remember having led the Norman-French army at the Battle of Hastings, this does not, in the estimation of most, make me the same person as William the Conqueror. On many accounts, I am not actually remembering anything (hence the scare quotes above). I think I remember leading the troops, but genuine memory is factive. To have any appeal at all, a memory theory will need to define identity in terms of actual memory and not apparent memory. The problem is that the difference between genuine memory and delusion seems to turn on a pre-existing fact about personal identity. I am deluded when I claim to remember William’s actions because I am not William. If genuine memory requires that the rememberer is the same person who had the experience remembered, however, a criterion of personal identity based on genuine memories is viciously circular.

Psychological continuity theorists address this concern by defining identity not in terms of memory, but in terms of “quasi-memory”. A quasi-memory is an experience as if of remembering (i.e., a representation of a past event with imagery and affect) which is “appropriately caused” by the experience represented. Here psychological theorists employ an influential view of memory defended by Martin and Deutscher (1966). They argue that what distinguishes remembering a past event from related phenomena like delusions, imaginings, or relearnings is that remembering involves a physical trace caused by the original experience and operative in bringing about the present representation of it. Psychological theorists then entertain science fiction scenarios in which memory traces can be copied or transplanted and argue that this makes it in principle possible for someone to have a quasi-memory of someone else’s experience which contains the elements relevant to constituting personal identity without presupposing personal identity. Defining identity in terms of quasi-memory rather than memory itself (with ordinary memory being a sub-species of quasi-memory)1, thus avoids circularity.

There are other modifications to the Lockean view in psychological continuity theories (e.g., adding other psychological connections besides memory to the criterion of personal identity, defining identity in terms of overlapping chains of psychological connections rather than direct connections) and many variations on this general approach. For present purposes, however, the introduction of quasi-memory is the most important, because it allows us to highlight the presupposi-

---

1 On this view, even memory-like experiences of our own past experiences do not count as memories unless they are also caused by the experience (rather than, e.g., by having heard the story of it so many times that one forms a vivid image of it). All apparent memories caused appropriately by the experience remembered are quasi-memories. “Memories” are just quasi-memories of experiences in the rememberer’s own personal history, and so a subset of quasi-memories.
tions about how memory contributes to the constitution of personal identity implied by these theories. It appears that according to psychological theorists, memory plays a role in constituting personal identity by (1) establishing similarity in the contents of consciousness over time (since the content of the representation in memory is taken to reproduce the original experience), and (2) providing a causal link between the rememberer and the experiencer. The contributions of the other kinds of psychological connections introduced by this view are seen to be of much the same kind. This picture is made explicit by David Lewis in “Survival and Identity”:

I find that what I mostly want in wanting survival is that my mental life should flow on. My present experiences, thoughts, beliefs, desires, and traits of character should have appropriate future successors. My total present mental state should be but one momentary stage in a continuing succession of mental states. These successive states should be interconnected in two ways. First, by bonds of similarity [...] Second, by bonds of lawful causal dependance. (1976, p. 17)

I mentioned earlier that Locke does not claim that personal identity is constituted by memory connections, but rather by “sameness of consciousness”. While many have found this idea intuitively compelling, sameness of consciousness is a rather mysterious notion and Locke does not provide a very clear criterion of identity over time. Psychological continuity theories can be seen as attempts to develop the idea behind Locke’s vague view into something more precise. What is interesting, however, is that once developed these theories have been subject to the objection that the relation of psychological continuity as they define it lacks the depth and practical significance that we attribute to personal identity. Given that Locke’s account of personal identity was devised precisely to explain and justify its practical significance (this is the idea of person as a “forensic” concept), the way in which psychological continuity theories are developed does not seem to capture what is compelling in Locke’s notion of “sameness of consciousness”.

Locke’s picture, I think, suggests a continuity of subjective self-experience throughout the life of a single person which provides a rich phenomenological connection to one’s own past and future, and a subjective sense of continuing in time. This is absent in the criterion provided by psychological continuity theorists. Instead, we are given appropriately caused chains of similarity in the contents of consciousness between distinct moments of temporally local experience. This deep experiential connection to one’s past and future, however, is arguably the very thing that gives identity its practical importance. If this is so, and we hope to give an account of identity that grounds the connections we draw between facts about personal identity and judgments of moral responsibility and the rationality of prudential concern, we will need to provide an understanding of sameness of consciousness that includes such an experiential connection. Derek Parfit, famously, argues that this cannot be done. Although we tend to think that personal
identity over time is a deep relationship, rife with practical significance, it is in fact
nothing more than what the psychological continuity theories implies, a superfi-
cial relation that cannot bear the practical weight put upon it. He suggests that we
should revise our practices associated with morality and self-concern in light of
this fact. (1986, pp. 245–347)

My suggestion is that revisiting the nature of memory and its potential contri-
bution to personal identity points to a strategy for avoiding Parfit’s extreme con-
clusion. The conception of memory employed in psychological continuity theories
remains essentially the one put forward by Martin and Deutscher in 1966. But a
great deal has happened in both the science and philosophy of memory since that
time. While these developments do not speak directly to this debate nor to mem-
ory’s role in personal identity, they do offer new ways of thinking about memory
and so point to a way of understanding how certain kinds of memories can provide
a deep, experiential connection to the past and, via projective imagination, to the
future. This relation, I suggest, can ground a connection between personal identity
and practical concerns. The view offered here is not a general account of personal
identity that will do all the work we would wish for such an account to do, but it
is one that can help us understand the appeal of Locke’s view (and the ubiquity of
the idea that memory constitutes personal identity) and show how it might fit into
a more general framework.

2 Recent developments concerning episodic
memory

There have been several recent developments in our understanding of memory
which, at the very least, suggest that episodic memory does not work in quite the
way that seems to be presupposed by psychological continuity theorists. To begin,
there is now virtual consensus among those studying memory that the episodic
memory system is highly reconstructive and rarely reproduces past experiences
with exact fidelity (see e.g., Addis et al., 2007; Bartlett & Burt, 1933; Loftus, 1979;
Neisser, 1981; Roediger et al., 2001; Schacter et al., 1995). Details of remembered
episodes might be altered at the time of encoding, during consolidation into long
term memory, and/or at the time of retrieval, after which they can be altered again
as they are reconsolidated. Lived experience alone demonstrates convincingly that
our memories very often turn out to deviate from their target episodes in at least
some respects. I may remember your dinner party last summer and how much
fun I had at it, but mistakenly recall that someone was present who was not, or
that it was at that dinner party where you first served your famous chicken when
it was really at a later dinner, or picture sitting around a rickety old table that
you had already replaced. Some philosophers of memory have reacted by pointing
out that if memory is rightly understood as a faculty for storing and faithfully
reproducing past experiences, it is a surprisingly bad one, deficient to an extent

©The author(s). https://philosophymindscience.org ISSN: 2699-0369
that naturally evolved systems rarely are. This raises the question of whether it is right to understand the purpose of episodic memory as faithful reproduction of past episodes, or to think that mismatches between the details of our memories and of the episodes remembered are necessarily deficits. Obviously, there is some sense in which memory is supposed to represent the past, and what is represented in memory cannot deviate too drastically or fundamentally from what occurred if we are to consider it successful. This does not necessarily imply, however, that the ideal of memory is always (or ever) to reproduce the past exactly and that failing to do so is therefore always a defect. The idea that episodic memory is not obviously a system for representing past episodes as faithfully as possible has gained considerable momentum over the past several years, for a variety of reasons (although see Craver, 2020 for a view that complicates these considerations).

One main line of argument picks up from the observation made earlier that one would not expect a naturally evolved system to go wrong as frequently as our memory appears to do if its purpose is faithful representation of past events. If we put aside our presuppositions and ask, from the standpoint of conferring survival advantage, what the purpose of episodic memory might be, it seems likely that it would be primarily to efficiently present information about how things have been in the past in a way that is useful for effectively guiding present and future action. Considered this way, exact fidelity is not necessarily optimal. It is rare that what we need to know to guide action is exactly what happened on some specific past occasion. It is quite possible, in fact, that we would usually do better to have a representation of the past that approximates the original episode but also updates and contextualizes its details in the light of knowledge gained in other relevantly similar circumstances.

A variety of empirical findings can be taken to suggest that this is indeed how memory works. While studies reveal that details of events are often altered in memory, they also reveal that these alterations are usually not haphazard, but systematic (see e.g., Intraub & Hoffman, 1992; Lindsay et al., 2004; Neisser, 1981; Neter & Waksberg, 1964; Nigro & Neisser, 1983; Payne et al., 1996; Roediger & McDermott, 1995). Often the altered memory represents the past in a way that it very probably might have been rather than the way it actually was. I might, for instance, remember a colleague who is almost always present at faculty meetings as having been present at a particular meeting where her presence or absence had no major effect, even though she was in fact absent that day. Memory representations are responsive to cues, prompts, and corrections from the environment, including the social environment, which may lead us to represent past events slightly differently from how they occurred. It is by no means obvious that these kinds of alterations in general make episodic memory less valuable to survival than exactly faithful reproduction of the target episode would be, and perhaps reason to think it makes them more so.

Neuroscientific evidence is also invoked to support rethinking the nature and aims of episodic memory, especially a range of studies suggesting that the same
neural pathways operative in episodic memory are also active in hypothetical thinking about the past (representing a past episode that might have happened but did not), and in projective imagination (imagining episodes that might occur in the future) (see e.g., Addis et al., 2007, 2009; Addis & Schacter, 2008; Schacter & Addis, 2007; Szpunar et al., 2007; Szpunar, 2010). Felipe De Brigard (2014) concludes from this evidence that remembering is “a particular operation of a cognitive system that permits the flexible recombination of different components of encoded traces into representations of possible past events that might or might not have occurred, in the service of constructing mental simulations of possible future events” (De Brigard, 2014, p. 158). He suggests, in an especially catchy slogan, that memory may “not be for remembering”, meaning that “many ordinary cases of misremembering should not be seen as instances of memory’s malfunction, but rather as the normal result of a larger cognitive system that performs a different function, and for which remembering is just one operation.” (De Brigard, 2014, p. 158) This general picture of how memory fits into our overall neurocognitive makeup is also at the center of Kourken Michaelian’s “simulationist” view of memory (2016). Focusing on similar considerations and kinds of empirical evidence, Michaelian argues that there is no need for representations of past events to have any causal connection to the events remembered to count as genuine memories. It may happen that none of the stored information used to reconstruct a representation of a particular episode was laid down during the experience of the episode itself. If the purpose of remembering is to guide present and future action, it does not matter whether the information out of which a present representation of a past action is constructed was encoded at the time of that episode or some other time at which similar information was obtained. An episodic memory on this view is just a representation of an episode generated by a properly functioning episodic scene construction system which is aiming to produce a representation of an event from one’s personal past. The provenance of the traces used in constructing that representation is irrelevant. Michaelian also has a catchy slogan: “remembering is imagining the past” (2016, p. 120). This slogan describes a less radical position than it might seem. The idea is not that we can have a genuine memory of any episode we can imagine. Michaelian is adamant about distinguishing between imaginings that are memories and “mere imaginings”. For an imagining of the past to be an episodic memory, it must result from the operations of a properly functioning episodic construction system aiming to represent an event from one’s personal past. For an episodic construction system to be functioning properly, its attempts to represent past episodes will need to usually issue in representations that are largely accurate to target episodes in one’s personal history. The relevant notion of “accuracy” here is accuracy subject to the kinds of systematic reconstructive revisions we have already discussed. A properly functioning system will, that is, produce, representations that can reliably guide us in our actions in a way that wild confabulations cannot. Episodic memory, projective imagination, and counterfactual thinking have come to be known collectively as “mental time travel” (MTT), and I will use this terminology in what
follows. Michaelian’s ultimate claim is that episodic memory is not a fundamental kind but should be understood as a sub-species of MTT.

The views described here give us some reason to question the simple picture of the role of memory in constituting personal identity that we found in psychological continuity theories. The deeply reconstructive nature of memory these accounts describe complicates a straightforward assumption that memories generate similarity of the contents of consciousness over time. The mechanisms by which reconstruction occurs, meanwhile, raise doubts about whether such memories can be relied upon to provide a causal connection to the past of the sort assumed in many psychological continuity theories. The results themselves are not uncontroversial, of course (see e.g., Robins, 2016 for objections), and I do not claim that they decisively show the standard view to be untenable. They do, however, point to a picture of episodic memory considerably more complicated than that assumed in psychological continuity theories and introduce the broader concept of MTT as a different way of thinking about our connection to our own past and future. It is worth considering whether this picture can provide a deeper unity within the life of a person than does the relation defined by psychological continuity theorists.

To show how it might, I begin with reflection on the broad epistemic claim that, contrary to the standard view, the aim of episodic memory might not be to faithfully reproduce past events, and so that memories can succeed in their aim without such faithful reproduction. Particularly important is the idea that deviations from faithful reproductions should not be seen as merely tolerable deficits but might in fact represent assets relative to memory’s actual purpose. In the context in which this claim is made, emphasis is on the evolution of the mechanism, and so the “aim of memory” is understood as maximizing chances of reproductive survival in the niche in which the capacity evolved. The idea that the success of a memory depends upon its aims can be generalized. Whatever aims might have been involved in generating the mechanisms of episodic memory, now that we have such mechanisms we can and do employ them for a wide variety of reasons.

Once we give up on the idea that faithful reproduction of past events or experiences is the only possible aim for memory, we can recognize that the other forms of MTT also have a variety of aims, and hence of success conditions and constraints. These may, moreover, overlap with or mirror those of relevantly similar forms of memory. Ultimately, my goal is to zero in on a type of MTT whose structure and constraints enable it to contribute to a phenomenologically rich connection to past and future that supports the practical judgments associated with personal identity. To meet this goal, it is useful to first think a bit more generally about the various aims MTT might have and the associated success conditions. I will start by thinking about this question with respect to imagination, since the features I wish to highlight are somewhat easier to see in this context.2

---

2 It is worth making explicit that this will involve applying the ideas I take from De Brigard and Michaelian well beyond the context in which they develop them, and possibly employing them in ways they would not endorse.
3 Imagination

One common way of distinguishing between memory and imagination is precisely through their success conditions. Memory, it is often said, is constrained by facts about the world, while imagination is not. Having a good memory is getting the facts about the past right, while having a good imagination is (at least partly) being able to bracket facts about the world and create something novel. This is what makes Michaelian’s slogan that remembering is imagining the past sound initially so jarring. As we have seen, the actual claim is not as radical as it first appears because the form of imagination Michaelian associates with memory is constrained by facts about the past. Imagining is only an instance of episodic memory if it is generated by a properly functioning episodic construction system with the aim of providing a relevantly accurate reproduction of an event in one’s autobiographical history.

The qualification of relevance is important since, as we have just seen, relevant accuracy is not always (or usually) perfectly faithful reproduction of all the details of the past, but rather of those which are relevant for memory to best meet its aims (in this case, aiding reproductive success). Once we understand that the constraints on accuracy required to make memory successful are indexed to purpose in this way, however, the idea that memory and imagination can be distinguished in this common way becomes somewhat tenuous. This is partly because memory, on this view, is less beholden to facts about the world than we might have thought, and partly because imagination can also have purposes that place accuracy constraints on it not dissimilar to those Michaelian sees as applying to episodic memory.

Amy Kind, a central figure in the philosophy of imagination, notes the heterogeneity of acts of imagination and develops an account of what she calls “imagination under constraints” (Kind, 2016). In some cases, she says imagination is not a mere flight of fancy. It is a method of simulation aimed at solving a problem. We might use imagination in planning or predicting in much the same way we might undertake a computer simulation to predict a storm path or traffic patterns. Kind offers the examples of Nikola Tesla, who reported that much of the engineering involved in the invention of alternating current technology was done via imaginative simulation rather than physical model building, and Temple Grandin, who greatly improved the design of livestock-handling facilities by imagining how different designs would look from an animal’s perspective.

Tesla and Grandin used their imaginations to develop effective interventions in the world. Meeting their aims thus required getting relevant facts about the world right. Tesla’s imaginings needed to be accurate to facts about electricity and materials, and Grandin’s to animal psychology, for their imaginative tasks to succeed. These are examples of individuals with extraordinary powers of imagination, but

3 Going forward, “accurate” should always be read as “relevantly accurate” in this sense, unless otherwise indicated.
this kind of exercise is quite common. We might, for instance, imagine a heavy piece of furniture at different spots in a room to determine which placement is most aesthetically pleasing or functional without having to move it. This everyday kind of imagining is similarly constrained by real-world facts. As Kind herself points out, however, imagination is heterogenous. There are many other purposes for which we imagine as well, with correspondingly different success conditions and so different kinds of accuracy constraints. I have no clear taxonomy of the purposes of imagination, and the distinctions here are undoubtedly messy, involving overlapping forms of imagination rather than tidy categories. Thinking about some examples of different uses to which imagination might be put, along with their success conditions and the corresponding accuracy constraints they impose will, however, be instructive, helping us, to appreciate the complex structure of a form of imagination closely associated with personal identity.

Sometimes when we imagine, our purpose is simply to amuse ourselves. If I am bored on a long train trip, I might make up a story just to pass the time. Here the imaginative exercise is successful if it entertains me. It is not obvious that such imagination needs to be constrained by facts about the real world to succeed. In some ways it may entertain best if it is outlandish and takes me out of my mundane circumstances, although there will likely need to be at least some familiar features in my imagined story to make it coherent enough to be entertaining. On other occasions, imagination has the somewhat more specific goal of evoking a particular experience or affect. There are various circumstances in which we might imagine in this way, and these will have differing success conditions and constraints. If I am feeling low, I might imagine a set of happy circumstances with the goal of lifting my spirits. If I am feeling insecure before an important presentation, I might imagine myself involved in brave and heroic adventures to feel more confident. If I am anxious or stressed, I might calm myself with a guided meditation in which I imagine myself in a peaceful wood, or on a tropical beach with a gentle breeze blowing.

When imagination is used in this way, its success consists in producing the desired affect and the constraints on successful imagining are those that allow it to do so. Undoubtedly this requires getting some facts about my own psychology right in the imaginative exercise. If I do not know what circumstances will induce the affect or experience I am seeking, the undertaking will not succeed. Crucially, however, there is no reason to think that in general the events imagined are or ever will be actual for this type of imagination to succeed. Imagining a spectacular tropical vacation might cheer me up even if I have no plans or resources to take such a vacation and do not expect to, and imagining slaying a dragon might fill me with confidence, even if I am pretty sure I will never encounter one (and would run away if I did).

In another form of imagination, however, I imagine a circumstance that I do expect to be part of my actual life, and I do so with the aim of evoking, in the present, the affective experience I expect these future events to engender. These are the
kinds of imaginative exercises I wish to connect to personal identity. An example of this sort of anticipatory imagination is found in *Macbeth*, where Lady Macbeth, having learned that her plan to become queen has been set in motion, reports that she feels “the future in the moment”. We might also think of Kevin McCarthy in January 2023 possibly sustaining himself through 14 painful defeats to be elected Speaker of the U.S. House of Representatives by imagining the moment when he would finally take the gavel, as he felt certain he would. Or consider a beleaguered faculty member getting herself through the last few hectic weeks of the term by imagining the fabulous vacation she has planned as soon as her grades are turned in. The success conditions of this form of imagination are complex. Like other forms of imagination in which we are trying to evoke affect, such acts of imagination will need to engender the affective state at which they are aiming. Unlike cases of “mere” imagination, however, success here seems to depend also on the events imagined being (relevantly) accurate to what happens. If the events imagined do not transpire, or if they do transpire but the affect they bring about is vastly different from what was anticipated, this form of imagination has not succeeded. Or so I argue.

The claim that the success of this form of imagining depends not only on evoking the desired experience but also on the accuracy of both the events anticipated and the affect they evoke when they occur, requires some explanation. The basic idea is that it is one thing to cheer yourself up in the depths of a frigid winter by daydreaming about a fabulous tropical vacation that you know full well you are not planning to take, and quite another to look forward to your upcoming actual vacation, with your tickets in hand, your hotel reserved, and your bags packed. One way of seeing this difference is to think about what happens in the latter case (but not the former) if things do not go as planned. If we are simply daydreaming, the hope is that the pleasant imagining will provide a bit of respite from our present dreary situation. The pleasure evoked by imagining what it will be like to enter those last grades and head for the airport as I actually expect to do, by contrast, depends upon the belief that the pleasure experienced now is a foretaste of pleasure that will be experienced in the future. That this pleasant experience will be in one’s future is part of what is generating pleasure now and part of what gives the imaginative exercise its potency in helping our professor make it through those last hectic weeks. This form of projective imagination is thus vulnerable to facts about the world. If the anticipated events do not come to fruition the pleasure

---

4 To keep the discussion manageable, I am focusing on circumstances in which imagination (and other forms of MTT) are actively undertaken with a conscious aim. I think the fundamentals of my analysis apply also to many involuntary, spontaneous forms of imagination and MTT. Relatedly, I am focusing on pleasant anticipations and memories. I think the connection to personal identity I am drawing applies also to those that are unpleasant – for instance if I am anxious about the future or regretful about the past. I will say a bit about this later but will not be able to explore these forms of MTT in detail here, or to defend the claim that they work in essentially the same way as those that are voluntary and pleasant.
quickly turns to disappointment. In this sense, the imaginative exercise has failed in a way that defeats its original aim.

Some might insist that the imaginative exercise itself has not failed in such cases, that if, for instance, imagining her upcoming vacation did cheer the beleaguered professor up and get her through the difficult weeks at the end of term, it has done its job. The fact that she is disappointed later might show that there can be a cost to such imagining, but not that it is deficient with respect to its own aims.

I see the pull of this view, and I am not sure that anything important turns on whether such acts of imagining are called defective or costly. I do think, however, that there is a way of thinking about why we might want to say the former that helps to illuminate the structure of such imaginings as well as their connection to personal identity. We might, I suggest, think about this kind of imagination as a form of “borrowing” affect from one’s future. In the U.S. there is a current television ad for a payday/structured settlement loan company with the tagline: “It’s your money, use it when you need it” (often followed by a potential customer shouting defiantly: “It’s my money, and I need it now.”) I might take out a payday loan in anticipation of my next paycheck with the aim of using now, when I need it, money that I have already earned and thus will be able to pay back later. If my boss absconds with the payroll and I never get my anticipated paycheck, however, something has gone wrong with my loan. I did get the money when I needed it, so that part of the aim was fulfilled, but I turned out not to have gotten it when I needed it and been in a position to pay it back later, and this matters.

Pleasurable anticipation is different from a payday loan in many important ways, of course. For one, it is not obvious that there is a determinate sum of pleasure to be spread out between now and the future as there is a determinate sum of money. When I “borrow” future pleasure I don’t expect to have to pay it back. Still, to the extent that the disappointment one experiences when an anticipated good is not forthcoming is deeper and more bitter than it would otherwise be, the analogy is not entirely out of place. Even if some disanalogy remains, moreover, there is a key structural similarity between the kind of anticipatory imagination of future pleasures and payday loans that is of particular importance for present purposes. Namely, the failure of the loan (when my boss leaves with the payroll) and of the anticipatory imagination (when my vacation gets cancelled) each involve my making a mistake about what is mine. In both cases, I am enjoying in the present a good that I take to be properly my own, even though I do not have immediate access to it. In the former case, the good is cash and in the latter, it is affect I expect to experience in the future. In both cases, however, there is a pre-

---

©The author(s). https://philosophymindscience.org ISSN: 2699-0369
supposition that the good is already mine (It’s your money, use it when you need it!), and this is why borrowing seems permissible. The possibility of failure exists because I am vulnerable to finding out that in the end what I thought was mine to use turns out not to be.

The kinds of anticipations in which we borrow affect from an anticipated future, and so which can fail in this way involve what I will call “strong identification” with a future person. In taking particular future pleasures (or pains) to be, in the sense described, already mine, and so experiencing them in anticipation, I am taking the person who will be subject to these pleasures (or pains) to be me. Her fortunes are my fortunes, and they matter to how things are for me right now. It is the strong form of identification found in these kinds of anticipations (and other forms of MTT) that I will connect to personal identity. Before saying more about how this works, however, it will be useful now to consider cases of memory with a similar structure.

4 Memory and counterfactuals

Having described the affective borrowing that characterizes some forms of prospective imagination, it is not difficult to see that we can and do borrow affect from the past in a corresponding way in some episodic memories. Once again, drawing examples from voluntary memories of events with a positive valence we can think about the way in which many people keep scrapbooks and mementoes so they can revisit and remember the good times, or couples play “their song” to recall the joy and excitement of when they were courting, or people remember past triumphs to psych themselves up for a present challenge, cheer themselves when they are lonely by remembering happy times with loved ones, or calm themselves when they are stressed by recalling peaceful vacations by the sea. If the analogy for using imagination to evoke affect in the present is borrowing against an anticipated payday, the analogy for memory would be tapping into savings, in this case from a bank of events or experiences in one’s personal past that carry with them a specific affective valance.

The possibility of this use of memory is widely recognized in everyday thought. It is common to hear memories described as “treasures”, which we are advised to “hold on to” so that we can have them when times are hard. This idea is expressed in jazz standards like Just the Way You Look Tonight (“Someday, when I’m awfully low/When the world is cold/I will feel a glow just thinking of you/And the way you look tonight”), in films and novels (e.g., in Casablanca, Rick and Ilsa decide as they part, probably forever, that they will “always have Paris”), on greeting cards, in poetry, and many other contexts. I have explored this form of memory in some detail in (Schechtman, 2022).

With memory, too, I think my analysis applies also to involuntary and negatively valanced examples but cannot make that case here.
Structurally, the cases in which we use memories to borrow affect from the past mirror the cases in which we borrow affect from the future, and their success conditions are similar. To succeed they need to evoke the intended affect and also to be accurate to (the relevant) facts about the world. These memories, like the corresponding forms of anticipation, involve strong identification, in this case with a past person. It is worth thinking a bit more explicitly about this kind of identification as it occurs in the context of memory, since it is similar to, but slightly different from, identification in the case of imagination.

The importance of the fact that these “treasured memories” cannot be taken away from us expressed in common wisdom, seems to lie not only in the fact that they give us pleasant things to think about, but more fundamentally in the fact that the events remembered are taken to belong to us. This is what distinguishes their role and impact from other forms of memory that do not involve strong identification. Memory with strong identification brings with it an experience of myself, right now, as the person who did those things and had those experiences. They are part of who I am and that can never change. Remembering that you gave your hometown its first-ever championship with your heroic game-winning play is not just revisiting a pleasant experience; it is reminding you of who you are and what kinds of self-regarding emotions you have the right to. Maybe you are, at this moment, being chewed out by the sales manager for once again failing to meet quota at the used car lot where you now work. But even if you are a mediocre used car salesman, you are not just that. You are also the hometown hero and remembering this fact can alter, in complex ways, the present experience of being dressed down by your manager.

Such memories are constrained by relevant facts about the world in the same way that anticipatory imagination with strong identification is. The affect evoked by these memories relies upon a belief that my representation of past events is accurate in the ways that matter to the quality and character of the affect that is their aim. If the used car salesman misremembers the jerseys worn during that game, the memory can still succeed. If it turns out that the play did not succeed and that his team lost, or that the team won but his part in the win was trivial, however, he may be crushed in a way that represents the cost of having borrowed affect from the memory all these years. Through strong identification he has tied his fortunes to a past person and so is vulnerable to finding out that he was mistaken about what those fortunes in fact were.

Although I will not discuss this point much further in this paper, there is a related form of counterfactual thinking about the past that involves strong identification. Consider “I coulda been a contender” as uttered by Marlon Brando in the film On the Waterfront. Thinking about the possibility that one did not agree to throw the fight and in consequence lived a far different life, is entertaining the assumption that there was a time when one had a genuine chance at greatness, experiencing both pride in the possibility and regret from having left it unfulfilled. This is a very different thing from someone with no particular musical talent or
charisma idly thinking about what it would have been like to be a rock star. The first kind of counterfactual thinking can engender self-regarding affect associated with the imagined possible events. It involves a commitment to the idea that it is true of me that this is a possibility I had but did not achieve (and one which might, in some cases, still be possible). This kind of counterfactual also depends upon facts about the world. For it to succeed, the world must be such that it is plausible to believe that had I made different choices I would have been a contender, that this was not just a pipe dream or a delusion. Again, strong identification brings with it a vulnerability to turning out to be wrong, finding out that I was never as good as I thought I was and would never have been a contender no matter what I did.

From the developments in the philosophy of memory described earlier, I took two ideas. One is that the success conditions and accuracy constraints on memories depend upon their aims. The other is a connection among memory, imagination, and counterfactual thought, which have been found to operate via similar mechanisms. Exploring and expanding upon these ideas, I have suggested that one aim of the various forms of mental time travel can be to evoke specific emotional or affective experiences. Sometimes this is done independent of any assumptions about the accuracy of what is represented. There seems, however, to be a specific kind of self-regarding affect that is evoked by representation of what are taken to be the details of our own past or future, or genuine possibilities left unfulfilled. States that aim at evoking affect in this way, I have said, involve strong identification, and are vulnerable to failure if they get the relevant facts wrong. This is the kind of complex state that could plausibly contribute to constituting personal identity by providing a deep experiential relation to temporally remote parts of one’s life that psychological continuity theories do not. I do not rule out that other forms of MTT might also play a contributing role. For now, however, I focus on a better understanding of exactly what kind of role these states with strong identification might play, and what kind of identity they might be connected with.

5 Mental time travel with strong identification and identity

This is perhaps the time to acknowledge that neither the role of MTT with strong identification in constituting identity I have described, nor the kind of identity it can help constitute, is simple or straightforward. The picture that emerges from the above analysis is immensely complex, and here I can do no more than describe its basic contours and gesture in the direction of further development. Perhaps the best way to start is with a relatively straightforward view which is not, in fact, the one I am proposing, but which might plausibly be seen to follow from what has been said.

Given the exposition so far, it would be natural to suppose that my claim is that we make past and future actions and experiences ours by remembering and

©The author(s). https://philosophymindscience.org ISSN: 2699-0369
anticipating them (for the sake of simplicity I will leave counterfactuals to the side for now) with strong identification. This basically takes up the memory theory standardly attributed to Locke but specifies a targeted set of memories (and anticipations) which are the ones that have identity-constituting powers. This view, however, would be subject to the same fatal flaws as a simple memory theory. That view, you will recall, faced a dilemma. If it says that any action or experience I seem to remember is mine, it is deeply implausible. If this implausibility is avoided by saying that I must seem to remember it and it must be something that I actually did or experienced, the view is viciously circular.

A view of identity based on strong identification in the straightforward way just described faces a similar problem. The fact that I strongly identify with making that game-winning play, and am filled with pride at having done so, does not seem to make me the person who did it. If it did, why could I not similarly identify in this strong way with leading the military victory at Hastings and make those actions mine? A view based on MTT with strong identification might seem to have an advantage here, since these states are said to succeed only if they are accurate to the relevant facts. This, however, threatens to land us on the other horn of the dilemma. If the relevant form of accuracy is accuracy to facts about personal identity, the view is viciously circular. And if it is not accuracy to facts about personal identity, it is not obvious it will solve the problem. A substitution of “MTT with strong identification” for “memory” in a simple memory theory thus faces serious difficulties, and this is not the view I am suggesting. Explaining how my proposed view avoids this dilemma will, however, allow me to clarify its contents.

To begin, it will help to get more precise about what one is identifying with in MTT with strong identification. What we remember or anticipate are actions, events, experiences, emotions, and the like – details of a time, place, or event. But what I have said we identify with is the person involved in these events, taking these actions, or having these experiences and emotions. But what, exactly, does it mean to identify with the person in question? So far, the only content I have given to this idea is implied by the accuracy requirement and the vulnerability it confers. If I imaginatively anticipate my upcoming vacation with strong identification, I am identifying myself with the person who is joyfully vacationing. If it turns out that that person is stuck in the office and must cancel the trip, I am committing myself to the likely disappointment and bleak mood she will experience. If I recall my glory days of football, I am identifying with that team member, and if it turns out his glory days were not as glorious as I remember, I may be devastated.

This does not yet tell us how I pick out and individuate the person with whom I’m identifying. Saying that I “identify with the future person who is on vacation” in anticipatory imagining is ambiguous. It could mean that I identify with whichever person it is who is on the imagined vacation, and so that the target of my identification is determined by who is taking that vacation at that time. Alternatively, it could mean that I identify with some independently defined future person who I anticipate will be on vacation at that time and take joy in the fact
that she will be. The first interpretation seems psychologically unrealistic and is, in any event, not the one I intend. If it is to be possible that I am mistaken in an instance of strong identification, and so vulnerable in the ways I have described to failing in my imaginative or memorial aims through mistakes about the world, I will need to pick out a person to whose fortunes I tie myself, and who I might be misrepresenting as myself. In practice, when we identify with a temporally remote person, we usually simultaneously represent actions and experiences we remember or anticipate them doing or having. The point I wish to emphasize is that the identification is with the person, which is why it affects us if we learn that things are not as we thought. This means that we can strongly identify with a future or past person even in an instance of MTT that fails on the grounds of inaccuracy.

Since the relevant form of identification is with the person and can take place even if I get the facts wrong, the view I am developing does not commit me to the view that any actions or experiences I remember are therefore mine, and so it does not suffer from that form of implausibility. This doesn’t fully solve the challenge based on William the Conqueror, however. In “remembering” my role in the victory at Hastings with strong identification, I have said, I am not identifying myself with “whoever in fact had that role at Hastings” but rather with a past person who I remember, perhaps wrongly, as having led the French-Norman army at Hastings. I might be extremely accurate about what happened at Hastings on that day in 1066 but be wrong that I (here, now) am the person involved in those events. I still need to make room for this kind of mistake, and it is not yet clear how I can do this. At this point it might not even be clear what “identifying with a person” amounts to. Whatever it means, it seems that to do it I will need to be able to individuate, identify, and reidentify persons. Moreover, if it makes sense to say that I might be mistaken in such an identification, as when I wrongly think I am William, there needs to be some fact about who I am which is independent of my strong identifications, in which case strong identification cannot constitute identity.

To explain how I intend to avoid this result I will have to give a quick and dirty description of the broader view I am trying to develop, of which MTT with strong identification is only one part. I have been saying throughout MTT with strong identification plays a role in constituting personal identity. On my view, however, it is not, the only element that plays such a role. It is part of what constitutes personal identity, but other elements must also be in place. The constraints placed by these other elements are what allow for the possibility of being mistaken in our identifications. Since these other elements also do not fully constitute identity on their own, however, there is still a crucial role for MTT with strong identification to play.

To get a sense of how this works and what some of these mysterious “other elements” are, it will be helpful to remember that the conception of personal identity we are exploring is essentially Lockean, in that it is tied to practical judgments and social relations Locke defends his view of personal identity on the grounds that it is
only sameness of consciousness that can support the judgements of moral responsibility and prudential concern we associate with identity. Locke’s own discussion and subsequent psychological accounts of identity have defined persons in terms of the cognitive requirements for being the kind of entity to whom such judgements apply (i.e., rational, self-conscious agents), and personal identity in terms of the identity of agents. The kind of strong agency essential to personhood on this view, however, requires an ability to interact effectively with the social and natural world, and this ability depends not only on facts about an individual’s psychological capacities, but also upon social and natural facts. It is indeed necessary to have a robust sense of oneself as a continuing self to engage in transactions with others and with the world, as Locke argues, but it is also necessary to have a reasonably accurate sense of social practices and the nature of the world around one, and this is an aspect of the forensic self which I believe Locke does not sufficiently emphasize and whose implications he fails to draw.

I thus claim that the conceptions of personhood and personal identity in play here require a person to be in general agreement with others about who is who and who did what, or at least to agree about how we determine and adjudicate facts like these. In our world, we individuate and reidentify people at the most basic level by individuating and reidentifying human animals. Our practices are all built around this. We know and acknowledge that people can change a great deal, possibly becoming almost unrecognizable, but when it comes to bank accounts, passports, baptisms, university degrees conferred, and crimes committed, what is in one’s human history is part of who one is. There can be disagreement about the significance or interpretation of events in a human history, e.g., I can remember myself fighting for justice and you can remember me breaching decorum, and this is something we can dispute. What we must agree on, though, barring any radical developments, is that if there is indisputable evidence that this human body was at such-and-such a place at such-and-such a time, then I was at that place at that time.

The fact that our practices are built this way is not a mere convention or accident of history. It is rooted in our form of embodiment and presumably runs deep in our evolutionary history. I do not know if it is necessary to arrange our practices in this way. There is no telling how technology might develop. In the world as it is now, however, persons are individuated and reidentified in this way. This has real consequences. If I have a crime in my human past, I may be prosecuted for it whether I remember it or not. If my human past involves the purchase of bonds which have since matured, the money is mine whether I recall buying the bonds or not and no matter how much I have changed. Because of this, we are socialized into thinking about ourselves as beings living a human life over time from earliest childhood. Children learn that there are consequences for past actions and are encouraged to anticipate traversing the various stages of human development (e.g., “soon you will be a big boy and go to school”, “one day you may have children of

---

7 I discuss this in detail in Schechtman (2014).
What I suggest, then, is that we are socialized into a shared conception of the constraints on legitimate strong identification as determined by the limits of a human life. What this means in practice is that in the typical case of MTT with strong identification, identification with a past or future person amounts to identifying with the past or future human being. If I remember Hastings in 1066, then, I am either representing this human, here now, as having been present and thus getting the details of its history very badly wrong, or I am identifying with the human being William the Conqueror and violating the shared understanding of the limits of individual persons that makes the social practices that support personhood possible. The inaccuracy of the identification thus depends partly on social factors, but not in a way that is arbitrarily constructivist. Without these practices, rooted deeply in our biology and facts about human development, there would not be persons as we know them at all.

At this point, it may no longer be clear what kind of constitutive role MTT with strong identity has left to play. If social norms determine the limits of what I can identify with, why not just say that they identify the limits of the person, independent of any remembering or anticipation on the part of the individual? The answer is that personhood, at least Lockean personhood, also requires a particular set of cognitive, agential, and affective capacities, and these depend upon an individual taking up the shared conception of personhood and personal identity and actively identifying with their human past and future. Unless I have a robust sense of myself as a continuing being, answerable for my human past and concerned about my human future, I will not be able to take my place in the social space of persons, interacting with the world and with others in person-defining ways. The kind of work that is done in MTT with strong identification, looking backward and forward and experiencing the vicissitudes of my past and future human self as mine, seems uniquely suited for engendering the kind of experience of oneself as a continuing self that is required for living and acting as a (Lockean) person.

If our goal is to provide an account of personal identity that tells us whether a person at time \( t_2 \) is numerically identical to a person at time \( t_1 \), MTT with strong identification will not, by itself, be able to do that. A fully formed criterion of personal identity of that form, if such a thing is possible, will require reference to social practice and perhaps much more. What my analysis of strong identification does seem to offer, however, is a promising way of getting at the kind of experientially rich connection to our past and future that “sameness of consciousness” seems to imply, and which psychological continuity theorists are unable to provide. A view of MTT in which we are affectively invested in the past and future so that their vicissitudes affect us in the present yields a phenomenology of self in

---

8 I leave to one side, for now, the question of the possibility of identifying with oneself in a different state after biological death. I do not see it as impossible, but the parameters are too complicated to discuss here.

©The author(s). https://philosophymindscience.org ISSN: 2699-0369
time that can make sense of the practical significance we place on personhood in a way that the view of memory as similarity plus appropriate cause cannot. I hope also to have provided some idea of how this piece might fit into the larger puzzle of how to think about the identities of beings like us.

Acknowledgments
I am grateful to participants at the online Workshop on Successful and Unsuccessful Remembering and Imagining, November 2022, and to Ying-Tung Lin, Chris McCarroll, and two anonymous referees for their extremely helpful comments.

References


Personal identity and mental time travel


Open Access
This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

©The author(s). https://philosophymindscience.org ISSN: 2699-0369