Michael McKenna

MORAL MONSTER OR RESPONSIBLE PERSON?
MEMENTO’S LEONARD AS A CASE STUDY IN DEFECTIVE AGENCY

CHRISTOPHER NOLAN’S psychological thriller Memento is an impressive cinematic achievement. Although many films can be used to illustrate philosophical ideas, few are intentionally philosophical. Memento is. In it, we are struck by several philosophical questions that the film is clearly designed to raise. Before proceeding to a discussion of these questions, I shall begin with a brief comment on the film itself.

As is often observed, Memento is nearly impenetrable on first viewing. Even a straightforward, unambiguous interpretation of the film, supposing one could find one, involves an extremely complex plot. Add to that the duplicitous innuendos and other built-in ambiguities, and we are provided with resources to doubt an unambiguous rendering. Finally, present the entire storyline using the intricate temporal structure Nolan developed, and the net effect of Memento is simply baffling, at least the first time through it.

If Nolan’s efforts were merely self-indulgent showmanship, it would be easy to dismiss this film. Intriguing as it is, it would be a fair criticism to complain that the film demands too much of its viewers, and that a finely crafted story in the fashion of Hitchcock or Bertolucci easily trumps all Nolan’s smoke and mirrors. But in my estimation, this complaint is misguided. Those pressing it fail to see the genius in what Nolan has accomplished. In terms of its complexity, I would argue that Nolan’s
Memento is similar to Faulkner’s masterpiece The Sound and the Fury. In both cases, the formal structure of the artwork—which no doubt completely overpowers its audience upon an initial encounter—provides a commentary on, or illustration of, the thematic subtext of the story. Only after revisiting the work can we come to see this, and in our revisiting we assign richer significance to the various pieces constituting the whole.

In Memento, Nolan has, in essence, demanded that his audience make use of mementos in order to follow his film. We, like Leonard, have to leave notes to ourselves to decipher the proper temporal order, think through the credibility of Leonard’s telling of the story of Sammy Jankis, scrutinize the veracity of Teddy’s revelations, understand the motives of characters such as Natalie, and so on. As we begin to question the stability of a single right interpretation of the film, we are led to reflect upon core philosophical themes in Memento, which turn on the nature of knowledge, mind, personal identity, and practical as well as distinctively moral agency.

**Philosophical issues in Memento**

Eventually I shall turn to an extended treatment of first practical and then moral agency. Here I wish to comment briefly on several other closely related issues.

**Global skepticism**

As the intelligent critic of Memento recognizes, Nolan has not given us enough textual evidence to settle what is to be taken for reality and what is not. Did Leonard kill his wife with an overdose of insulin, as Teddy alleges, or was it the assailant Leonard is seeking? How much of what we see is just a hallucination or, more innocuously, maybe a daydream, unfolding inside the confines of Leonard’s mind? Furthermore, how are we to sort through all of this, given the elaborate temporal ordering Nolan creates? Because of our uncertainty, we are placed in a position similar to Leonard’s. Thus, our relation to the film, like Leonard’s relation to the world he confronts, leaves us as Descartes found himself at the beginning of his second meditation, before he was able to convince himself that there was an external world. In this moment, he was uncertain as to how to interpret the status of his beliefs. He wrote:
So serious are the doubts into which I have been thrown as a result of yesterday’s meditation that I can neither put them out of my mind nor see any way of resolving them. It feels as if I have fallen unexpectedly into a deep whirlpool which tumbles me around so that I can neither stand on the bottom nor swim up to the top.

(1641: 16)

If we take all of the ambiguous interpretive data seriously, the effect Nolan creates with Memento is analogous to that created by unresolved Cartesian global skepticism. Given the available evidence, there simply is no settled, proper interpretation of reality. As between competing and inconsistent interpretations, we are completely ill-equipped to settle on which, if any, is veridical.

Descartes attempted to argue his way out of his skeptical predicament, and as we cast about for anchors in the film to rule out some evidence and retain other pieces, we are attempting to do likewise. We start with certainties, the ones closest to home and least subject to doubt. (Descartes claimed to find his first certainty in the mere fact of his own thought.) With this in mind, recall the scene in which Leonard has a meltdown in front of Natalie, whereupon, in a moment of tenderness, she takes him to bed (P). In that scene, Leonard insists that there are certainties, such as knowing what it will feel like when he picks up an ashtray. The practical predicament for him, as cast in that scene, is whether there are enough crumbs of certainty scattered about to build a bridge to his past, and to his future. For Descartes, his solipsistic predicament was whether he had enough to build a bridge to knowledge outside of his own mind.

But is Leonard really involved in this struggle while sitting in Natalie’s living room? Or is he just insane, imagining the entire thing from an asylum? Just as Descartes in his early stage of skeptical doubt could interpret all of his experiences as consistent with veridical judgments about the world, or instead as fabrications fed to him by a demon, so we in watching the film can interpret what Nolan presents for us as consistent with Leonard actually out hunting down a killer, or instead as delusional daydreams fed to him by his own insanity. The uncertainty that Nolan engenders does a far better job of illustrating the Cartesian skeptical predicament than a film such as the Wachowski brothers’ The Matrix, in which the audience is easily clued into which interpretation is reality and which is an artifice.
Self-knowledge, and knowledge of other minds

A striking fact about the human condition, at least in the normal case, is that there is a crucial asymmetry between knowledge of our own minds and knowledge of others’ minds. Admittedly, there is room for self-deception. Sometimes, maybe even often, people do not understand themselves nearly as well as they think they do. But the default assumption is that persons have a privileged relationship to their own beliefs about themselves. No doubt, the degree of credence that we ought to assign to their convictions varies depending on the beliefs in question. We regard people as completely authoritative about sincere pain reports, slightly less so about their visual impressions, less than that about what they claim to see, and so on. Once we get to things such as sexual preferences and how people really feel about their mothers, well, all bets are off. But along this spectrum, awareness of one’s own basic biography, at least in simple details, such as what one had for breakfast earlier in the day, or whether one cut the grass yesterday, are certainly regarded as beliefs about which one has a greater authoritative claim in comparison with third parties. My knowledge about what I had for breakfast this morning is more reliable than my knowledge of what you had for breakfast this morning. In the case of knowledge of my own history, I can consult my own “memory-record” of my experiences. But I do not have access to your experiences, to your memory-record. I have to infer knowledge of what you did without direct access to your mind.

The classical philosophical problem of knowledge of other minds is posed as a restricted kind of skepticism. Set aside the skeptical challenge that we do not know that there is an external world. Suppose that problem is solved; we do know there is an external world. A distinct problem is how we know there are other minds, and even if we do, how do we know the content of those other minds? Since we cannot have another’s thoughts or feelings, cannot experience her pains, and so on, all we can do is infer from her external behavior that she is having thoughts and experiences. But maybe she is not. Maybe she is a hollow zombie who just acts like we do when we have experiences. Or maybe she has them but, unlike in our case, there is no correlation between her conduct and her thought, so that no matter how she behaves, we are forever ignorant about what she is really feeling and thinking.

Although Leonard does not entertain full throttle skeptical doubts about knowledge of other minds, his relation to his own recent self is
just like his relation to another person. He has to infer what he had previously thought without access to any memories that would give him a privileged basis for any beliefs he might form. In this way, Leonard stands diachronically to his own recent self as he does synchronically to another person. The primary Socratic edict is to know thyself. Now recall Teddy’s skeptical challenge to Leonard, “You don’t even know who you are” (H, 1:07:52). For Leonard, knowing himself will always involve gigantic obstacles, ones that will make it nearly impossible for him ever to figure out what to do, since this requires some appreciation of what he has done (a point I will explore in more detail later).

The extended mind

What constitutes a mind? Put differently, what constitutes a subject of mental states or processes? Many philosophers, myself included, are physicalists. We believe that minds are physically realized. There is a good bit of technical philosophical machinery required to get clear on what physicalism does and does not commit one to. Setting all of that out is far beyond the scope of this essay, but a central idea is something like this: This mind of mine, Michael McKenna’s mind, whatever it comes to, is dependent upon physical states and processes, and is, in some loose sense of “is,” a physical state or process. So, roughly, my mind is located where my brain is located, inside my cranium. As peculiar as it might seem, saying this much, committing to physicalism, does not amount to the stronger thesis that my mind is just a physical state or process, where the “just” suggests the idea that mental states and processes can be reduced to physical states and processes. To explain: Though I contend that my mind is a physical state or process, and that, for instance, my beliefs, desires, intentions, values, and so forth, are all physical, it does not follow that they are simply physical, or that they are “no more than” physical. It might be that, though they are physical, what makes them mental states is that they function in certain ways. So if my mind were, say, built out of some different material, or if a small part of my brain quit working and could be replaced by a silicon chip that would perform the functional tasks now performed by the operative grey matter in my brain, it would still be my mind. Hence, while my brain does duty for how in fact my mind is “realized,” it is possible (even if not now technologically realistic) that it could be realized by some other physical material.
Nevertheless, on this view, when we point to my brain we point, for all practical purposes, to my mind. Stick my brain in a blender, reducing it to a frosty shake, and you do the same to my mind.

Given the assumption of physicalism, here is a striking puzzle brought out by *Memento*: If a mind is, in some sense, a physical system, then why assume that it is located only in the cranium, or under the skin? (Clark and Chalmers 1998) Why can’t it be extended outside of the skin too, for example, my computer, which retains articles I have written but have long since “forgotten?” Why is it not also extended to my weekly planner? What about snapshots of my father, long dead now? If this photo on my desk of him and me as a kid walking in the park sustains a memory of my time with him, why is it not part of my mind rather than something that can affect my mind? If you were to destroy it, my computer, or my daily planner, would I be entitled to say that you destroyed a part of my mind?

If the mind extends beyond the skin, then Leonard’s mementos are part of his mind. In fact, to the extent that he relies upon others to aid him in his planning (as Teddy has aided him), his mind is, in part, constituted by those others, just as if my daily planner is part of my mind, so too is my wife, who helps me to remember when to have the car scheduled for an oil change, or to decide what investments are most prudent, and so on. Alluding to his memory condition, at one point Teddy says to Leonard that he (Leonard) does not know anything. But assuming a theory of the extended mind, Leonard does know various things. It is just that he cannot use memories to get to them. Nevertheless he has access via other routes, and, as Leonard points out in arguing against Teddy, memories are overrated anyway. They are not records of the facts, he protests, but interpretations that can be deceiving.

**Personal Identity**

What makes a person the very same person that she is at a particular time and over time? By “the very same” I mean numerical identity. When we say things like “She is not the same person she used to be” we (almost) never mean that there was this one person, Cindy, but Cindy no longer exists, that is, that very entity—that person—is gone, and in her place is a completely numerically distinct person, Cindy*. What we mean is something like, one single person, Cindy, has changed so much that she.
that same person Cindy, has acquired very different qualities from the ones she used to have. She used to be unkind and selfish, but she—that same person—is now generous and sensitive. This kind of change is known as qualitative change, and is philosophically unproblematic (or at least less problematic). But numerical change, or rather numerical sameness—that is, lack of numerical change over time—is much more challenging. How do we track a being’s alterations across time such that, though that being does undergo qualitative changes, it remains numerically the same one? We can ask this question about any object, such as a car or a bicycle, as well as a person. If I buy replacement parts for each piece of my bicycle and have them all in boxes right by my bike, replacing just one part, like my handlebars, do I now have a numerically distinct bike? Of course not! It is my same bike with new handlebars. But what if I replace every part, one right after the other, in two hours? The bike before me after two hours shares no part with the earlier bike. Do I now have a numerically distinct bike? Of course! What could possibly explain these different answers? This illustrates the problem of explaining numerical identity over time.

When applying the problem of numerical identity over time to persons, philosophers have proffered various theories to account for how the same person can change radically over time and be the same person. One theory is the bodily continuity theory. If some person, say Leonard, is a particular human body, then he remains the same person across time so long as that body exists (albeit always changing). In Memento, on the bodily continuity theory, there is no problem about Leonard’s identity over time. The same animal body does all of these different things, so Leonard does them. But like many other philosophers, I find the simple bodily continuity theory highly implausible. Suppose that a demon were to completely erase all of my memories and psychological states overnight and replace them in this body with the memories and other states of the actress Betty Davis. I would regard this as an identity-destroying change. Michael McKenna would no longer exist, and some very confused man would be in his place. Or consider a real-life case, one that bears on memory-retention. Many family members of Alzheimer’s patients claim that, in the last stages of the disease, when a person’s memories are wiped clean and there is no memory of the person who was, “Grandpa has already gone; he’s not there anymore.” I think these claims should be taken literally. The person suffered an identity destroying change at some
point, and this person, or this human shell, is no longer that person, is no longer Grandpa.³

Reflections like these have led many philosophers to endorse some version of a psychological continuity theory of personal identity. A person remains the same person over time if and only if there is some sort of "identity grounding" relation between her psychological states across time. The philosopher John Locke first proposed memory as the crucial requirement (1689: bk. 2, ch. 27). And if we develop Locke's suggestion so that (most of) the memories have to be genuine (and not false), then we can link a person's identity across time like we link shorter strands of twine together to form a longer rope. I might have no memory now of my experiences when I was five years old, but I do have memories of my twenty-year-old self, and this self had memories of my ten-year-old self, and so on.

On a psychological continuity theory, Leonard's situation appears to raise deep metaphysical problems for him, aside from his gigantic practical problems. If Leonard in his current state has no recent memories of his recent past, then is he no longer the person who was assaulted, the one who was an insurance agent, and who was married? In one scene, Nolan raises this question when he has Teddy say to Leonard that he does not even know who he is, that he used to be Leonard, but is not any longer (H. 1:08.00). Here, it seems to me that Teddy is confused (maybe Nolan is too). Leonard has memories of his former self (before his accident), and so even on a psychological continuity theory, he is that same person. That he does not have memories of what he did an hour ago is of no matter, since he does have genuine memories that link him to his earlier self. No version of a psychological continuity theory could be true if it set the bar for identity over time so high that at each moment in time, a person at later times had to be linked to a genuine memory of that earlier time. Every dreamless sleep would result in a death, as would a serious alcohol bender, or a dance with the magic mushroom fairy.⁴

So, if Memento is to pose a puzzle about personal identity, it cannot quite be the one that Teddy (or Nolan?) thinks it is. The problem must be a more subtle one, and I think that there is one to be seen in the film. The key is to realize that a good theory of personal identity will not just account for how a person at one time is the same person she was in the past. It must explain what will be required for her to be the same person
in the future. In the normal case, this will allow a person some stable practical basis for caring about her future self. The problem for Leonard is not the fear that he will not be in the future. He knows that, unless he dies, he will be, and he knows that the ground for his being the same person in the future is that his future self will have access through his memories to the same history as he has now in his present state (access to his memories before his accident). Leonard’s problem is that, as he is now, he knows that his future self will not be able to know the present part of his biography. He knows that he will (probably) have a future, but he also knows that his future self will forever be a stranger to its recent biography, that is, to him as he is now. Thus, he cannot act now in the normal ways one might to help shape one’s future self, so that one might be able to grow or mature as a person. And this makes it especially hard for him to care about what he does now in terms of how it will bear on who he will be or how he will think of himself later. This is brilliantly illustrated in his fleeting, clear-eyed decision to set himself on Teddy.

**Practical agency**

I now turn to the issue of practical agency. As will become clear, the previous philosophical considerations will have some bearing on how we understand this topic, as well as the more narrowly focused topic of moral agency (to be discussed in the following section).

What is an action? What distinguishes it from a mere event in the world? Like many philosophers of action, I endorse a liberal view of what makes an event an action. Anytime a being’s behavior can be understood as purposeful by attributing to it even the most rudimentary beliefs and desires, we have action. This allows for actions by agents that are very far down on the zoological pecking order. If a bird is searching on the ground and moves off the pavement into the dirt, we can see his behavior as purposive to the extent that we can see it as minimally rational. The bird believes that the worms are in the dirt and not the pavement, and he desires some food. This explains his act of pecking.

But the actions characteristic of human agency are certainly more complex, and it is a difficult question just what is distinctive about the practical agency of comparatively sophisticated creatures like us. An especially challenging problem is to explain what makes an action
intentional. On a simple view—one that I think is indefensible—for each act, there is an intending that goes with it, a particular kind of mental state or event. Earlier today I drove to the gym. According to the simple view, if this was an act, then I intended to do it. So far, so good. But on the way to the gym I had to use my turn signals five times. On the simple view, if each use was an act, then I intended to do each. That seems a stretch, but not too outrageous. However, note that each time I directed the car in any new direction, I had to move the steering wheel ever so slightly. Each of these were surely actions of mine, but did I intend to do each? Was there a corresponding intention—a discrete mental state or event—for each of these acts of steering wheel manipulation? The view begins to sound implausible, and it is.

**Intentional action and planning agents**

Here is a rough sketch of a better view, due to the philosopher Michael Bratman (1987). Not every intentional action is one that a person intends. What makes an action intentional is that it fits within a broader plan, and what practical agents like us intend is to execute certain plans. So, in my action of driving to the gym, I intended to drive to the gym. When I used my turn signals and turned the steering wheel several times, each of those acts I did intentionally, but I did not intend them. All I intended in executing them was to drive to the gym. They were done intentionally by virtue of their figuring in how it was that I planned to get to the gym. Thus, our intentions are plans, and they can themselves be embedded in larger, and ever more complex plans. (Going to the gym is embedded in the larger plan of remaining fit, or training for a bicycling race.) What we do intentionally is to perform various acts in an effort to execute our sometimes immensely elaborate plans. Some of what we do intentionally is what we intend to do (e.g., go to the gym), but the majority of what we do intentionally (e.g., turn the steering wheel) is not what we directly intend. We only do these latter things (intentionally) in an effort to achieve what we do intend.

Well-functioning practical agents like us have elaborate legislative and executive resources for forming and carrying out our plans. We must, since our plans require on-the-spot modifications and adjustments, reassessment of the reasonableness of goals, and so on. New information can get us to revise how we execute a plan, or get us to reconsider whether
we should sustain a commitment to a plan in the first place. If on the way to the gym I learn that one route is closed due to road construction, I must be able to alter the details of my plan in order to sustain my commitment to the larger goal. If it is announced on the radio as I am driving that the gym is closed today, then I need to be able to abort the mission and revise my goals, and so on. Furthermore, this plan might need to be coordinated with others that could be assigned a higher or lower priority. Perhaps I also have a plan to get to the doctor's office, and if on the way to the gym I learn that I can only do both by rearranging the time I will be at the gym, then I will likely revise accordingly. On a planning theory of intentions, our practical agency is seen to be distinctive of creatures like us due to our ability to apply sophisticated patterns of practical reasoning to develop, sustain, coordinate, and carry out elaborate plans. All the while, our procedures need to be sufficiently plastic that we are able to remold them on a dime, based on the possibility of constantly changing bits of information, newly emerging preferences, reconsidered values that speak to the desirability of our goals, and so on.

Because, on this theory, intentions figure in a wider diachronic picture of an agent's goals, intentional action is best understood as intelligible only insofar as it is spread out over time. If so, it is easy to see that this makes problems for Leonard. The mementos Leonard uses are meant to facilitate his planning as best as he is able. And the different types of mementos that he leaves for himself are categorized into varying levels of importance—the most important of which he has tattooed on himself. Leonard's entire elaborate process, including his reliance on habituation, is all part of his seriously handicapped effort to sustain and coordinate his intentions, and embed them in plans that outstrip the limits of his memory's impoverished temporal horizon.

To call into relief just how impaired Leonard's practical agency is, Nolan has the hotel clerk, Burt, remark that Leonard's problem is the reverse of most people's (U, 0.09–15). Most people know what they have done, but don't know what they want to do. (In Bratman's terms, they have yet to settle on their plans.) Leonard seems to know what he wants to do next, but doesn't know what he just did. Recall also that at one point Natalie comments on Leonard's sorry system for coordinating his plans. She says that it must be hard to live your life "according to a couple of scraps of paper. You mix your laundry list with your grocery list and you'll end up eating your underwear for breakfast" (S, 0:18.04).
Because of Leonard’s condition, the prospects for exercising his agency are starkly limited. In a normal human life, the possibilities for meaningful action are varied, and they offer the prospect of many different rewarding paths into the future. Not so for Leonard. Think about his agency in terms of his relation to the future. The only way, it seems, that he can coordinate his agency so as to embed his immediate intentional conduct in a more expansive diachronic plan is by relying upon some single, simple thing to give his activity a purpose, an organizing goal around which the details of his immediate actions could make sense. This is what Teddy has done for him, sweet guy that he is. Now think about Leonard’s agency in relation to the past. Our present plans for how to act are shaped by our understanding of our recent history. It is on the basis of the recent events in our lives that we figure out what to do next. Our diachronic agency not only stretches forward in time, but reaches toward the future in light of an understanding of what the immediate past has made salient. If, indeed, the last thing he remembers is his wife dying, then intelligible options for Leonard will forever be from that moment forward. There are, therefore, very few options for him, other than to consider some act to aid in grieving (like hiring a hooker to help him say goodbye), or some way of seeking revenge. One can only wonder what will be left for Leonard once he offs Teddy. Maybe Leonard will be able to sustain a more expansive diachronic plan with no more than his mementos to guide him. But it seems very likely that, without Teddy’s competent mind aiding him, he’ll quickly crash and burn.

**Truth, desire, and self-deception**

So Leonard’s abilities as a planning agent are severely impaired, and this is due in large part to the fact that he is ignorant of so much. But that does not alter the fact that he still has a strong desire to regard himself as appropriately anchored in the world. This calls attention to a further feature of competent planning agents like ourselves. We set goals for ourselves—make plans—in light of our beliefs about how attempts to achieve those goals will fit with the facts of an independent world. This is partially for the instrumental reason that otherwise we’ll fail to achieve what we set out to do. But it is also because the goals we set for ourselves are often based upon reasons that presuppose values and a reality
independent of our preferences. Leonard, for instance, rejects the relevance of Natalie’s suggestion that his quest for vengeance is futile since he’ll never be able to remember achieving it (scene 5). It won’t matter to him, she tells him. But he’ll have none of this. It doesn’t matter if he won’t know, if he will be unable to reap the enduring satisfaction of being able to recall his righteous vengeance. He says that it still matters that the guilty man, John G (whoever he is), pay for killing his wife. She deserves to be avenged. Leonard’s desire that his actions be anchored in the truth is also revealed when he reacts with panic to the fear that he might have just killed an innocent man (Jimmy Grantz). In this moment, Teddy tries to reason with Leonard just as Natalie did: “So you lie to yourself to be happy—there’s nothing wrong with that!” (A, 1:42:28).

Thus, we see in Memento an illustration of the value of truth for a planning agent, even one whose relation to it is so very fragile. This aspect of planning agency is, however, often pitted against another, against desire. We form plans in order to achieve goals so as to make the world conform to what we wish. As it is often explained, one distinguishing feature between belief, whose goal is truth, and desire, whose goal is satisfaction, is the proper relation of fit between each and the world. In the case of belief, it aims to fit the world. In the case of desire, it aims to make the world fit it. Thus, these two ingredients, both necessary constituents of intentional agency, are uncomfortable bedfellows, often at odds with each other.

Nolan illustrates this tension with Leonard’s struggle. His desire for revenge has outstripped what is true, and at one point Teddy accuses Leonard of not wanting genuine truth, but of fabricating it so that he can live a dream of carrying out his campaign to avenge his wife. So furious is Leonard upon hearing this, and learning of Teddy’s manipulation of him, that he intentionally deceives himself, setting himself up to find and kill Teddy. In this moment, Leonard forsakes his allegiance to the truth, as he cynically says to himself, “Do I lie to myself to be happy? In your case Teddy, yes I will” (A, 1:48:15).

What Nolan offers us in the case of Leonard is either a tragic or a comic magnification of our predicament as planning agents, as practical agents trying to live in a world that is not our making. Our desires are sometimes at odds with our allegiance to the truth. Still, needing the latter to achieve the former, we are forced to constrain desire accordingly.
Nevertheless, sometimes the incentive for self-deception is extremely strong, so strong that we judge it worth the cost of forsaking truth, as Leonard does in that fleeting moment of clarity when he decides to give Teddy a taste of his own medicine.

It may well be that some degree of self-deception in human agency is a good thing. We are probably all a lot better off believing that we are a little better looking, smarter, more interesting, kinder, and more thoughtful than we really are. Were we to live in the constant light of the unvarnished truth about ourselves, we'd likely be so burdened that it would be hard to get up in the morning. All the same, as the case of Leonard illustrates, it is also clear that the limits of healthy or harmless self-deception are fairly constrained.

**Meaningful lives and narrative structure**

The fact that, as planning agents, our practical agency is spread out across time also helps to make sense of how we evaluate the meaningfulness of a life. By and large, we presume that a meaningful life is one that involves the pursuit and achievement of goals that sustain one over the relatively large span of a life. The relative worth of a life is a matter of its unfolding history, whether it is lived in pursuit of a reputable body of published work, an artistic oeuvre, or a career in a profession, trade, or even sport. So, too, for raising a family, investing one's life in the love of a spouse, or cultivating lasting friendships that color and enrich life. How such things unfold over time will speak to the value or meaning they offer to a life lived. A career that starts gloriously and then crashes on the rocks due to embezzlement and professional infidelity will be regarded as wasted not just for the later bad acts, but because the promise shown in its infancy was squandered. A career starting out from meager beginnings, with little hope for success, that then takes off and is sustained through many difficulties will be regarded, by contrast, as well-lived not just for the later successes, but because the earlier failures made the later successes all the sweeter. Those successes will be understood as part of a larger struggle to overcome adversity. Hence, the value of our lives, their meaning, for good or ill, is partially characterized by their narrative structure.

As is obvious, the prospects are very limited for Leonard's life having any meaning that would involve sustained narrative development. For
this reason, so many of the things that make a human life meaningful are beyond Leonard's reach. Consider grieving. Grieving involves a process of coming to accept and heal from the pain of loss. It can also be the source of future growth and possibility. The loss of a husband or a wife in an accident or due to an illness, while a source of pain, is also the source of the enriching possibilities that a future human life can offer, in the possibility of a new love, or a life lived in homage to the one who has passed. Because Leonard cannot forget, he cannot grieve, and so this distinctive pattern of healing and self-discovery is unavailable to him. As noted above, the same applies to the desire for vengeance, which for Leonard can never be quenched. Even if we find it an ugly, barbaric, primitive desire, part of its role in the unfolding of the life of the avenger is to facilitate a process of resolving the outrage of having been deeply aggrieved.

Even the one meaningful task Leonard seeks to achieve, one that would have narrative structure—the hunting down and killing of his wife's murderer—is undermined as something that could be meaningful for Leonard once achieved. Suppose he does leave evidence to himself that he succeeded, with a tattoo reading "I've done it" (Figure 1). He is nevertheless easily able to wonder if he had been deceived, and so did not really do it. Thus, concerns about the truthfulness of his mementos are likely to rob him of the little bit of meaningful narrative development his life could have.

Figure 1  Space for Leonard's final tattoo (12, 1:22:17)
Moral agency: moral monster or responsible person?

Is Leonard a moral agent? When he murders the drug dealer Jimmy Grantz, or later Teddy, or the various other John G’s that he has presumably killed, is Leonard morally responsible for what he does? Or is he a moral monster, a being capable of moral evil, but not really a candidate for competent responsible agency, perhaps like Doctor Frankenstein’s creation?

The foregoing treatment of the philosophical issues leading up to this topic offers a considerable part of the answer, or so it seems. It is reasonable to assume that morally responsible agency presupposes competent practical agency. And as I have been at pains to bring out, Leonard’s capacities for competent agency are severely restricted. March down through each topic: Leonard is living a life like the one Descartes described in the midst of global skeptical doubt. His knowledge of his own recent self is no better than his knowledge of the minds of others. Even assuming his mementos serve as part of his mind, on an extended-mind view, then so do Teddy’s machinations. And while his personal identity per se is not threatened by his memory condition, his ability to care about his current relation to his future self is hamstrung. Furthermore, as a planning agent he is highly limited, and to the extent that his actions can be coordinated around a long-term plan, he’s at the mercy of people like Teddy and Natalie. His commitment to the truth about the world—whether his wife is really avenged or not—is one he cannot reliably confirm, and he has a strong incentive to deceive himself, which eventually he does. Finally, the prospects for his life ever unfolding meaningfully in a way that could have the sort of narrative structure that others naturally value is beyond his reach. How could a person so off the radar, so impaired in his practical agency, be considered a competent moral agent, one who is responsible in any way for his actions?

Even if all of the above hurdles could be overcome so that we could still take seriously the claim that, in his very acts of murder and violence, Leonard is morally responsible for them, there is the further hurdle that we could never see Leonard as a candidate for our blame or punishment, nor could he even see himself as a candidate for guilt. The philosopher David Hume presented a famous puzzle about the gap between an act and an agent (1740: bk. 2, pt. 3, §2). When we blame and punish, we are led to do so because of the act. But if the act is over, then why punish
the agent? Hume's answer is that the act flowed from the bad character traits of the agent, and so we punish the agent as a way of responding to these traits. But not surprisingly, an agent who simply could not remember his acts could not see the blame or the punishment as responding to the character traits in him as they were revealed in his actions.

Still, it might be insisted, though we cannot ever meaningfully blame or punish Leonard, it does not follow that, in those very moments of his violent murders, he is not blameworthy, that he is not morally responsible for what he does then and there. True, we cannot blame him since he cannot remember, but we also cannot blame a man who murders and then drops dead immediately thereafter. Though dead, it does not follow that in murdering, the man was not worthy of blame. If he were to have lived, then it would have been perfectly fitting to punish him. So too, one might say, Leonard, in acting as he does, impaired as he is in various ways, is blameworthy, and were he able to remember, it would be perfectly fitting to blame him. The mere fact that we are impeded from blaming him does not detract from the credibility of the judgment that in acting he was worthy of blame.

Can we take this possibility seriously? Surprisingly, some philosophers endorse an ahistorical theory of morally responsible agency that would make good sense of how we could think of Leonard as morally responsible for his actions at the very times that he acts. They allow for the possibility that a person could come to be in a particular condition in a very bizarre way, one that subverted her agency and autonomy (that is, her self-government) altogether, but if, in the moment of acting, her agency works properly with respect to the here-and-now decision and action undertaken, then she can be morally responsible for what she does. Consider this famous quote from the philosopher Harry Frankfurt:

[T]o the extent that a person identifies with the springs of his actions, he takes responsibility for those actions and acquires moral responsibility for them; moreover, the questions of how the actions and his identifications with their springs are caused are irrelevant to the questions of whether he performs the actions freely or is morally responsible for performing them.

(1988: 54)
Leonard fully identified with the springs of his actions, and though he came to acquire them in deviant ways, according to the ahistorical theory defended by Frankfurt, that is of no importance as to the question of whether Leonard is morally responsible for what he did. When he acted, his mental life was, in the here-and-now sense, largely cohesive, organized as it was around a commitment to a campaign of revenge. And while his reasons for killing, in the case of Jimmy Grantz, for instance, were the products of others’ acts of deception, nevertheless he still acted on them. Furthermore, he was well aware, as he notes several times, and as others note to him, that his evidence was fairly shaky for who did what and which clues he could rely upon. Nevertheless, on the basis of this very thin evidence, he was all too prepared to kill. These sorts of in-the-moment decisions reveal deeply morally objectionable ways in which Leonard exercised his agency. And so, why not, as Frankfurt’s thesis would have it, respond to these aspects of Leonard’s agency as grounds for judging that, indeed, he was morally responsible for what he did at those different moments of action?

Numerous other philosophers writing on morally responsible agency defend an historical account of moral responsibility, arguing that an agent’s history matters as to whether she is morally responsible for what she does at that time. Indeed, they would argue that if you were to imagine some other character, Leonard*, who had no memory problem, who was not hoodwinked by others into believing all of these things about Jimmy, but who had come to a decision to murder Jimmy out of revenge under his own rationally competent steam, then Leonard* would be morally responsible. Leonard, however, would not, even though Leonard and Leonard* would be just alike in all of their ahistorical (or “current time-slice”) properties at the moment of action. At best, they would argue, Leonard is a moral monster, but surely not a morally responsible agent, and so not morally responsible for his actions.

I confess that between the historicists and the ahistoricists, I have remained ambivalent about who is right. I see the force of Frankfurt’s thesis that we evaluate people as they are, regardless of how they came to be that way. The poor kid from nowhere and the wealthy kid from privilege, once they have hit a certain threshold of competent agency, should be evaluated just based on who they are and what they do now, not how they came to be that way. On the other hand, it stretches imagination to think that, in the moment of action, we could really regard
Leonard to be just as blameworthy as Leonard*. In an earlier publication (McKenna 2004), I floated a way to try to make more plausible the idea that someone like Leonard is just as responsible for what he does as someone like Leonard* would be. I argued that a character like Leonard*, unlike Leonard, is morally responsible for more than the mere fact of his action. He is also morally responsible for the character that he freely formed on his own, the one that led to his coming to be the guy who, just like Leonard, performed such and such act of murder. Thus, we can make sense of our thought that Leonard* deserves more blame than Leonard deserves. On the current proposal, it is not because Leonard deserves nothing and Leonard* deserves to be blamed for the bad act of murder. Rather, both deserved to be blamed for the bad act of murder, but only one, Leonard*, deserves to be blamed for being the person he is and for coming to be the person who performed this bad act.

Thinking carefully about the case of Leonard has actually led me to be less comfortable about my previous proposal. If Leonard is morally responsible and blameworthy for his acts of murder, and so is not a moral monster but, in those moments, a morally responsible agent, his moral responsibility is so vanishingly minimal that it seems not to amount to much. If Leonard could be morally responsible for his acts of murder despite the numerous and colossal impairments to his states of knowledge, self-understanding, and practical agency that I have enumerated in the preceding discussion, then one wonders how much responsibility he could bear, how much it would be reasonable to say that he shared with a Leonard* who by contrast is a largely normally functioning human person. Very little, I think. Very little indeed.

A fine film, like a fine novel, if it is of the right sort, can provide especially vivid thought experiments designed to test philosophical intuitions. I close by reporting that because of Memento, I have been led to revise my earlier commitment to a defense of the ahistorical thesis. Even if there is any life in it, the moral responsibility that could be attributed to an agent like Leonard strongly suggests that it’s not worth the ink spilt to discuss it.7

Notes

1 In what follows, I will assume familiarity with the film, and I will operate under the assumption that the film does create interpretive challenges that
make any unambiguous rendering of it contentious. I should also note that I have profited from Andy Klein’s impressive essay “Everything You Wanted to Know about Memento,” (2001) although I am not in complete agreement with it.

2 In this way, Memento is like Brian De Palma’s Body Double (1984). Chris Grau has also recommended David Cronenberg’s eXistenZ (1999) and Joseph Rusnak’s The Thirteenth Floor (1999).

3 For an opposing view, see Williams (1970).

4 For another film that explores interesting questions of personal identity, see Paul Verhoeven’s Total Recall (1990).

5 On this topic, I have profited from Hibbs (2003). I am indebted to Chris Grau for calling my attention to this essay.

6 The points considered in the preceding paragraph are owed to Matthew Talbert’s judicious comments.

7 I am grateful to Christopher Grau for his thoughtful and extensive comments on an earlier draft of this paper. Matthew Talbert, acting as a reader for Routledge, also offered excellent suggestions, and so I thank him here. Finally, I would like to thank Andrew Kania for inviting me to contribute to this volume, and for his wise and patient editorial advice.

References


Further reading


