Is Remembering to do a Special Kind of Memory?

Grünbaum, Thor; Kyllingsbæk, Søren

Published in:
Review of Philosophy and Psychology

DOI:
10.1007/s13164-020-00479-5

Publication date:
2020

Document version
Peer reviewed version

Final draft. Accepted 2020 in Review of Philosophy and Psychology.
Please quote the published version: https://doi.org/10.1007/s13164-020-00479-5

Is Remembering to do a Special Kind of Memory?

Thor Grünbaum (Section for Philosophy, University of Copenhagen)

Søren Kyllingsbæk (Department of Psychology, University of Copenhagen)

Abstract

When a person decides to do something in the future, she forms an intention and her intention persists. Philosophers have thought about the rational requirement that an agent's intention persists until its execution. But philosophers have neglected to think about the causal memory mechanisms that could enable this kind of persistence and its role in rational long-term agency. Our aim of this paper is to fill this gap by arguing that memory for intention is a specific kind of memory. We do this by evaluating and rejecting standard declarative accounts of memory for intention and arguing for the plausibility of an alternative model of memory for intention. We argue for the alternative by spelling out a number of computational principles that could enable retaining and retrieving intentions from long-term memory. These principles could explain a number of core features of intentions.

1. Introduction

I lie in bed in the evening and decide to tell a colleague about the new coffee machine when seeing her in the morning. Between the evening and meeting at work in the morning, I do not give it any thought. I recall my intention to tell her, when I see her in the hallway. When recalling my intention to tell her, I do not reconsider my decision. I do not need to make up my mind again and re-endorse the action. When recalling my intention, I simply consciously intend to tell her about the new coffee machine.

In this paper, we will assume that we sometimes make effective decisions like this. This type of description could be true for some of our delayed actions, but by no means for all of them. A number of important implications follow from this basic picture of temporally extended human agency. Some of these are implications concerning the causal memory mechanisms that enable this kind of future-directed decision-making. First, recalling an intention involves some form of selection among a plurality of relevant intentions retained in long-term memory. Agents make decisions about future actions continuously throughout the day. Consequently, at any given time, many intentions are

retained in long-term memory. Selecting the right intention at the right time from long-term memory is crucial for rational temporally extended agency. Second, recalling one's intention is often almost automatic. Given the context, often one's intention simply pops into mind. Third, the motivational profile of an intention (the agent's commitment and action-readiness) is also retained in long-term memory. When bringing my intention to mind in the morning, my intention comes to mind as an intention, i.e. with the motivational force of an intention.

In this paper, we will argue that the long-term memory involved in retaining one's intentions over time and bringing them to mind again, when the agent needs it, is a special kind of memory. This kind of memory for intention is distinct from episodic and semantic forms of declarative memory. Memory for intention has a distinct type of content, function, and cognitive dynamics that enables it to play a distinctive role in rational temporally extended agency. We argue that episodic and semantic kinds of declarative memory are unable to play this role. To substantiate the claim that episodic and semantic accounts are not the only possible accounts of memory, we sketch a computational theory according to which intentions are represented in long-term memory by information relevant for attention, motor execution, and the rational role of the intention in a larger goal-hierarchy. Our aim is to present a "how-possible-account" of the computational mechanisms that might enable a specific type of memory for intentions. Together these lines of argument give us reasons to accept that memory for intention is a special kind of declarative of memory.

Many philosophers have written about decision-making and the nature of intentions but, curiously, only few have directly addressed how agents retain and later recall their decisions and intentions and the role of this kind of memory in temporally extended agency. The aim of this paper is to fill this gap by providing an account of representations of intentions in long-term memory and the possible mechanisms retrieval of intentions into working memory. We spell out a number of computational principles that could enable retaining and retrieving intentions from long-term memory. These principles can explain a number of core features that would characterize intentions if a Bratman inspired view of intentions were true.

We will proceed as follows. In Section 2, we sketch a Bratman inspired account of future-directed intentions. In Section 3, we use this account to derive three explanatory desiderata for a theory of memory for intention. In Section 4, we discuss and reject two families of semantic and episodic accounts of memory for intention. In Section 5, we outline a number computational principle that could provide us with a possible account of memory for intention. According to these principles,

memory for intention is a special kind of memory. In Section 6, we argue that the account can satisfy the desiderata.

2. Future-directed intentions

When an agent decides to do something, she forms an intention to do it. An agent can now decide to do something now and thereby form a *present-directed intention*, or she can now decide to do something in the future and thereby form a *future-directed intention*. Given the limited nature of human working memory, processing capacity, and processing speed, it is impossible for agents to compute optimal utility functions on the spot in everyday situations of action. One aspect of this problem concerns the fact that the psychological process of deliberation and decision-making takes up time and cognitive resources – resources that in a given situation might be better used on monitoring the environment. For a capacity limited human agent, it will be crucial that the time of deliberation and decision-making can be separated from the time of action. That is, it is crucial that she has the capacity to form, retain, and later recall future-directed intentions.

Future-directed decision making is governed by an important rationality constraint. It is rational for a capacity and processing limited agent to engage in this kind of future directed practical deliberation and decision-making, only if intentions can remain with the agent over time. If intentions lost their power over time, were too easily reconsidered, or always required reconsideration and re-endorsement at the time of action, it would be irrational for the agent to engage in future-directed deliberation and decision-making. If all intentional actions were decided immediately before the moment of action, one ought to not bother spending one's time and cognitive resources on future-directed deliberation and decision-making (Bratman, 1987).

According to Bratman (1987), human future-directed decision-making is rational, only if the following three features characterize the resulting future-directed intentions. First, intentions are "conduct-controlling". When the time and context for the action is present, the intention will directly control the agent's behaviour. That is, when the time for action comes, the agent will at least try to execute her intention. By contrast, strong desires and wishes to do something will merely raise the probability of an agent doing the thing (by raising the probability that she decides and thereby forms the intention).

Second, intentions have "stability". There is a cost involved in reconsidering one's intentions, and sometimes it is costlier for an agent to reconsider her intention than simply to stick with it. After deciding and forming an intention to do something tomorrow, the agent will start making other plans

conditionally on the intention. Revising one's decision is likely to start a cascade of planning adjustments. Such a revision of one's intentions will not only have intra-personal costs; often there will also be social costs. Social coordination is often dependent on the stability of intentions. By contrast, desires do not seem to have the same kind of stability. Even my strongest desires – like my desire to look at my social media updates on my phone – might co-exist with a constant disposition to reconsider whether to act on the desire.

Third, intentions play an important role in practical reasoning and human rationality. Not only are an agent's intentions constrained by her beliefs, her intentions also feed into her practical reasoning. Intentions are often schematic in the sense that they do not specify all the details of a future performance. In this way, such intentions generate a disposition to engage in further reasoning and place constraints on such reasoning. Furthermore, this later deliberation and decision-making must be consistent with the agent's other intentions and background knowledge. In playing this rational role, intentions differ from desires. It is not irrational to have conflicting desires (I might have both the desire to finish a given task without interruption and the desire to look at my phone now – and they might both be equally strong). Moreover, it is not irrational to have a strong desire to do something and decide to do something else instead.

Summing up, future-directed intentions are characterized by three functional features: conduct-control, stability over time, and rational role in reasoning. Memory obviously plays important roles in enabling these functions. If an agent is unable to recall her intention to do something, her intention will not be able to control her action, exhibit stability, and play a rational role in reasoning and decision-making. For the rest of the paper, we will assume that a plausible theory of memory of intentions, at both encoding, storing, and retrieval stages, should in part provide an explanation of the mechanism that enable these three functional features of future-directed intentions.[1]

One important implication seems to follow from this picture of future-directed intentions: A person should be able to retain intentions in long-term memory at a very low cognitive cost to her on-going reasoning and action. Let us see why. A person is able to represent multiple intentions in her memory. In the morning, Ann decides to phone her sister in the evening after dinner. A few minutes later she sees today's TV-programme and decides to watch *Monkey Business* beginning sometime after having usually finished dinner. Later, just before lunch, she realizes that she will have to take

[1] Note that by making this assumption we are assuming that Bratman's (1987) account of future-directed intentions is largely correct. Consequently, intentions are assumed to be qualitatively different from desires, and what we say about memory for intentions should not be assumed to transfer to an account of the role of desires in temporally extended agency.

some work with her home and decides to finish preparing the lecture late in the evening. She now has many future-directed intentions.

Multiple intentions kept active in a capacity limited form of working memory for a considerable stretch of time will be detrimental to one's ongoing task performance. Rather than using up the limited capacity by keeping multiple intentions present to her mind, it would be better for the agent to have a long-term memory mechanism. A human agent with the ability to represent multiple intentions in long-term memory for later retrieval should be able to represent them at a low cognitive cost. This in turn requires us to be able to make sense of the situation where it might be true to say of an agent that she intends to φ even if this intention is not present to the agent's conscious mind or otherwise operative in the agent's reasoning and decision-making.

One way to make sense of the claim that a person is in some psychological state even if the agent is not consciously in this state is to distinguish between *standing* and *occurrent* psychological states. Take the case of beliefs. I might acquire a belief by reading about some fact in a newspaper. I might come to think that President Donald Trump wants to buy Greenland. After reading the newspaper, I might become engaged in entirely different matters, I might even take a nap, and all the while, it would still be correct to say of me that I believe that Donald Trump wants to buy Greenland. During this time (or at least during some of it), the belief is merely standing. It plays no operative role in my mental life. In the evening, when discussing the latest developments in international politics with a friend, I recall that Trump wants to buy Greenland. The belief becomes occurrent, i.e. it comes to play an active role in my thinking.

The distinction between *standing* and *occurrent* is related to cognitive psychologists' distinction between long-term memory and working memory, respectively. Working memory processes are constrained by *specific* capacity limitations (on average about four items). By contrast, long-term memory has no *specific* capacity limitation (there seems to be always room for one more item). Understanding the notions of standing and occurrent in this way enables us to see that conscious awareness might be just one way for a state to be occurrent. More generally, a state is occurrent if it is directly involved in cognitively capacity limited activities such as reasoning, decision-making, and problem-solving. A state is standing if it is not currently directly operative in any such working memory demanding processes. Following the view of dominant computational processing models of memory (Anderson, 1983; Cowan, 1993, 2019; Fuster & Alexander, 1971), we assume that the difference between long-term and working memory should be understood in terms of activation levels. Whether content is standing in long-term memory or occurrent in working memory is simply

a question of its level of activation. We do not need to posit two separate systems of memory and imagine that content is sent from one system to the other. This activation level conception of the difference between content in long-term and working memory will become important as we proceed.

One way to make sense of standing beliefs is by a distinction between *dispositional belief* and *dispositions to believe* (Audi, 1994). When engaged in other matters or when taking a nap, I have a standing belief about Trump and the buying of Greenland, which is disposed to become occurrent (given the right retrieval cues). Dispositional beliefs are in this sense different from a mere tendency to acquire beliefs in particular circumstances. Without having read about Trump in the newspaper this morning, I might nevertheless be disposed to acquire the belief that Trump wants to buy Greenland, given all the other stuff I already know about him, Greenland, and the geo-political importance of the Arctic regions. It is not always easy to distinguish between dispositional beliefs and dispositions to belief. One important feature of dispositional beliefs (i.e. standing beliefs) is that they stand in the right causal relation to a corresponding prior occurrent belief.

Plausibly, a similar story could be told about desires and intentions.[2] In the morning, when Ann is forming the intention to phone her sister in the evening, the future-directed intention is occurrent in her working memory, and she is using it as an anchor to plan her evening. However, later in morning, it is no longer occurrent in her working memory but is standing in long-term memory as a disposition to become operative at a later time. When a future-directed intention is operative in decision-making, means-end reasoning, problem-solving, or action control, it can be said to be *occurrent*. When the intention is not cognitively operative (but it is still true to say of the agent that she intends to do the action) it is *standing*.

A standing intention is a special kind of disposition to form an occurrent intention. However, not all dispositions to form occurrent intentions are standing intentions. An agent might simply be disposed given her other intentions, desires, and beliefs, to form a new occurrent intention to φ when perceiving the right cue. This is not the same as having a standing intention to φ. Standing intentions are a special kind of dispositions to form occurrent intentions – the kind where the dispositional intention has the right causal relation to a prior occurrent intention (i.e., before being standing, it was occurrent).

In accordance with the activation level conception of the difference between long-term memory and working memory, we will understand the difference between standing and occurrent intentions

[2] Concerning standing desires, see Alston (1967, p. 402) and Goldman (1970, pp. 86-88). For a discussion of these sources, see Mele (2003, pp. 30-33). Concerning standing intentions, see Mele, 2007.

as a difference in activation levels. Consequently, we should not understand an occurrent intention and a standing intention as separate entities where one cause the other. Rather, it is the same representational entity in different states of activation. Following this view, in this paper we will use the notions "retrieval" and "recall" interchangeably, as simply denoting the processes by which a standing intention becomes active and ready for use in working memory processes (reasoning, planning, decision-making, and action control). If one accepts these claims, then providing an account of the memory mechanisms that enable the persistence of intentions includes a theory about the factors that explain the changes in activation.

3. Explanatory desiderata

Accepting Bratman's three features of future-directed intention, the distinction between standing and occurrent intentions, and the processing view of activation levels has interesting consequences for theories of retrieval of standing intentions from long-term memory.

First, the mechanism responsible for retrieving standing intentions from long-term memory (thereby making them occurrent) includes a selection mechanism. Often some time will elapse between the decision to do an action and the execution of the action. In between the two, the agent will often be occupied with other cognitive activities. When the time comes, the agent has to retrieve her future-directed intention from long-term memory and transform it from a standing to an occurrent intention. The problem is that at any time there will always be a multitude of standing intentions represented in long-term memory. So, by which mechanism is the agent able to select the right future-directed intention from this multitude? The retrieval mechanism must be able to solve the problem of how to select and activate an appropriate intention in a given context.

One might worry that from the claim that long-term memory represents multiple intentions it does not follow that the agent or her cognitive system is confronted with a selection problem in any interesting sense. One could imagine that all intentions are encoded and retained in memory with a time indicator or number. Intentions are stored and released as a sequence by their number: intention_{30965}, intention_{30966}, etc. Or one could imagine all representations of intentions in long-term memory were encoded analogously to a stack of discs where the top disc is the active representation.[3]

It is instructive to consider why models like these are wrong. To work, the agent should be able to have near perfect foresight. The agent should be able to plan accurately for one possible future, and plan for this future by explicitly planning for one action after the other. That is obviously not the

[3] See Altman & Trafton (2002) for a discussion of "goal stack" models.

situation for normal human agents. We have to plan for an uncertain and opaque future. We often plan for several possible futures. Consequently, many of our intentions are conditional on other actions and events obtaining. And furthermore, often our planning is schematic, only to be filled in with further planning at a later stage. As a result, many of our intentions are often tied to similar cues in the surroundings, they are often overlapping, and occasionally they are competing. A rational agent therefore needs a cognitive system that can select the right intention, given the current sequence of actions and cues in the environment.

Second, retrieval of standing intentions from long-term memory is "automatic" in the sense that it is usually not a conscious, voluntary search activity. Agents often rely on environmental cues to trigger the retrieval (McDaniel & Einstein, 2007; McDaniel & Einstein, 2000). The cue might be consciously perceived but the triggered process of retrieving the intention does not require any conscious steps. Often, the intention simply pops up in one's mind (Kvavilashvili & Mandler, 2004; Reese & Cherry, 2002). This does not make the retrieval a completely passive process (let us therefore call the retrieval process "semi-automatic"). The agent can influence the process in various ways, for instance, by strengthening the intention at encoding (e.g. by rehearsing it) or creating external reminders (post-it notes, etc.).

While agents might not deliberately control the retrieval of their standing intentions from long-term memory, the execution of their intentions is another matter. The fact that standing intentions often become active in working memory automatically does not imply that intentions are automatically executed.[4] It is the retrieval of the intention that is triggered by environmental cues, not the execution of the action. There are usually several cognitive steps involved before an agent engages in the execution of an action. Sometimes an intention becomes active in working memory because the agent needs to figure out how to perform it; sometimes the intention becomes active because the agent needs to check new possible plans against standing commitments; sometimes it becomes active because the agent needs to monitor the environment for cues signalling the right time to execute the action (Guynn, 2003). Consequently, the present discussion is about the selection of standing intentions for working memory and not directly for execution.

Third, the process of retrieving a standing intention in long-term memory and making it active in working memory does not need to involve new deliberation and decision-making. This follows from Bratman's functional features of conduct-control and stability. When a standing intention becomes occurrent again, it comes online with its motivational force – as an intention. Often,

[4] For empirical support of this point, see Hommel & Wiers, 2017.

retrieving does not involve a process of making a new decision to perform an action. Often, retrieving a standing intention can simply make the intention occurrent, and reconsideration and acts of recommitment are unnecessary. By contrast, forming a new intention usually requires that an agent makes up her mind and commit to the action.

Summing up, we should expect that human agents capable of temporally extended agency are able to retain multiple standing intentions at a very low cognitive cost. Given the Bratman picture of intentions, the distinction between occurrent and standing intentions, and some plausible computational assumptions about cognitive processing, we have reasons to accept three explanatory desiderata. Theories of the mechanisms of the memory for intention should be able to explain (1) how the retrieval mechanism solves the problem of selecting an appropriate intention from the multitude of potentially relevant standing intentions, (2) the often semi-automatic character of the retrieval, and (3) how standing intentions can become occurrent again without the need for any re-endorsement or new decision-making.[5]

4. Semantic and episodic theories of memory for intention

To our knowledge, there exist no detailed philosophical theories of memory for intention. The few we have identified assume that memory for intention is a familiar kind of declarative memory. In this section, we will put the explanatory desiderata to work in the evaluation of two types of account of memory for intention. This should give us reasons to accept that memory for intention is neither a semantic nor episodic kind of memory. In the next section, we will argue that it could be a special kind of declarative memory.

By a "declarative memory", we mean a cognitive form of memory where stored contents cause behaviour only in various indirect ways by entering into working memory related processes such as reasoning and decision-making (Schacter & Tulving, 1994). The standard taxonomy is that declarative memory consists of separate episodic and semantic systems of memory (Squire, 2004; Tulving, 2002). By "episodic memory", we mean a kind of memory that makes reference to events experienced in one's past and to one's experience of the events in one's past. By "semantic memory", we mean a kind of memory where one remembers a fact, that may or may not be about oneself, but where one does not necessarily know the circumstances under which one acquired the information.

[5] It is by no means implied that these are the only important cognitive features of memory for intention. For instance, as in other types of memory, there might be important constructive processes involved in retrieval and roles for metacognitive assessment in accepting or rejecting a recalled intention. Here we focus on the features that are more closely associated with Bratman's account of intentions.

Our claim is that we should add a third kind of memory to the standard taxonomy of declarative memory.

In this section, we consider two accounts of memory for intention according to which memory for intention is either a semantic or episodic kind of memory of making decisions or having reasons. These accounts of memory for intention are abstract and schematic. Plausibly, they could be specified and supplemented with computational principles such that they would satisfy the desiderata (1) and (2). In Section 5, we will state a number of computational principles that would make a retrieval mechanism satisfy (1) and (2). There is no reason to suppose that the accounts considered in this section could not avail themselves of similar principles. In this section, we will focus on desideratum (3), that standing intentions can become occurrent again without the need for any re-endorsement or new decision-making.

4.1. Semantic or episodic memory of decision-making

One might think that there is no need to postulate a special kind of declarative memory for intention. Semantic or episodic kinds can do the job just fine. What it means to say that an agent can recall her intention to do something is simply that she can recall a relevant fact about herself – this could be the fact that she made the decision or at some prior time had an occurrent intention – and that this causes her to form a corresponding occurrent intention.

We can picture the course of events like this. An agent decides in the evening to tell her colleague about the new coffee machine when seeing her in the morning. This causes her to form the occurrent intention to tell the colleague in the morning and to form the occurrent belief that she made the decision and intends to tell the colleague something in the morning. This occurrent belief then becomes a standing belief. Given the right cues (e.g. seeing the colleague in the hallway), the agent retrieves her standing belief. That is, she recalls that she decided in the evening to tell the colleague something in the morning. The now occurrent belief about her prior decision causes her to now occurrently intend to tell the colleague. According to this proposal then, there really are no long-term persisting intentions, though it is still be true that there are dispositions to form occurrent intentions, and these dispositions are caused by prior occurrent intentions. There is a standing belief that is caused by an occurrent intention and that goes on to cause an occurrent intention.[6]

[6] This sketch is similar to Mele's proposal for a sufficient condition for standing intentions (Mele, 2007, p. 750). Paul (2012) also seems to endorse a similar proposal.

If memory for intentions were this retrospective form of memory about past facts concerning oneself, then it would be either an episodic form of memory or a semantic form of memory. Recalling that some event occurred is an episodic form of memory if the person recalls the event's having occurred, where this entails bringing to mind also certain aspects of experiencing the event. By contrast, recalling that some event occurred is a semantic form of memory if the person recalls that the event occurred and she does not recall the event's having occurred.[7] She might remember simply because someone had told her about it, though note that it is not necessary that she has knowledge about the source. Accordingly, memory for intentions is either a form of episodic memory, semantic memory, or a combination of both.

Memory for intention could not be a semantic form of memory. A person would be semantically remembering her prior intention if she is remembering the intention she had when she was a young child but where this knowledge comes not from having experienced the episode but from being told about it (perhaps without having any awareness of the source). She remembers her intention but not because she remembers the episode of deciding or intending having occurred. This possibility would seem to introduce a kind of epistemic openness that does not seem possible with remembering one's intention. When semantically remembering that I intended to φ, it would make sense to wonder "but was it really me who intended to φ?". I could get it right that someone had made a decision but be wrong about who it was. On the basis of my semantic memory, I could be correctly thinking that someone intended to φ but could at the same time be wrong that the person was me. But self-awareness of occurrent intentions seems to be immune to this kind of error (O'Brien, 2003; 2007, Part II). If it does not make sense for an agent to be aware of intending to φ but be confused about whether it is herself that so intends to φ, then being aware that someone intended to φ on the basis of semantic memory could not be the same as intending to φ. An extra cognitive step of identification and endorsement is needed before one's awareness that someone intended to φ becomes one's awareness that oneself intended to φ and one's occurrent intention to φ now.

In contrast to semantically recalling that I intended to φ, when I recall my intention to φ, my intention becomes active as an occurrent intention, and there is no room for doubts about ownership. If there were, I would not be recalling my intention in a form where it becomes active. This points to a general problem with a semantic memory account of memory for intention. It is possible to semantically recall that I decided or intended to φ without now being committed to φ-ing. That is, semantically recalling one's intention is possible without the intention becoming occurrent. To make

[7] For a related way of drawing the distinction, see Naylor, 2011.

the intention occurrent, the agent would need to make a new decision or recommit to the old intention. This would violate the third desideratum for explanations of memory for intention.

If remembering intentions could not be a semantic kind of memory, could it be an episodic kind? After all, episodic memory is often thought to exclude a related kind of epistemic openness.[8] The point is this. Just as a person's first-person judgements about her own occurrent intentions seems to be immune to error about who is the subject of the intention, so first-person judgements on the basis of episodic memory appear to be immune to error about who was the subject of the remembered experience (Shoemaker, 1970; Fernández, 2014). If that is the case, then an episodic memory account of memory for intention would not be vulnerable to the foregoing problem with a semantic memory account.

There are at least two reasons for thinking that recalling one's intention is not a form of episodic memory of one's prior decision. First, episodic memory is essentially past directed, whereas recalling intentions is future directed. As a form of episodic memory, I recall the event of having made the decision or the episode of having been committed. I recall when and where of making the decision or being committed. By contrast, when I recall my intention, I make occurrent my commitment to do something. In making this standing commitment occurrent, I orient myself to the action to be performed. When recalling my intention in this sense, it is not necessary that I recall the past event of making the decision. When seeing my colleague in the morning, I come to intend to tell her about the coffee machine. I need not mentally travel back to the event the previous evening of making this decision this decision to tell her, though something like this ability might be necessary for a rational agent's ability to give reasons for her action.

It could be objected that many psychologists emphasise the future-oriented function of episodic memory. The notion of memory for the future is now commonplace in psychological studies of episodic memory (Atance and O'Neill, 2001; Klein, 2013; Schacter and Addis, 2009). Thus, it might be false to claim that episodic memory is essentially past directed. This objection is confusing facts about the adaptive function of the capacity for episodic memory and its underlying neural machinery with facts about the intentional objects of the exercise of the capacity. What I recall is the taking place of some event in the past. In recalling the event, I mentally relive aspects of the past experience. Even if the mind is thus directed at a past experience, the purpose of such a capacity might be to enable the agent to mentally simulate possible situations of perception and action (Boyer, 2008).

[8] See, for instance, Fernández (2006, p. 43): "Once I appear to be remembering a certain event episodically, the question whether or not I seem to have perceived that event is no longer open."

Second, as in the case of semantic memory, the episodic memory account of memory for intention introduces an extra cognitive step where none ought to be. A cue prompts the agent to retrieve her standing belief about a past episode of deciding to φ (or a time in the past when she was consciously intending to φ), and this now occurrent belief causes her to occurently intend to φ. But such beliefs do not always translate seamlessly into corresponding intentions. Recalling that I decided to φ or that I intended to φ does not guarantee that I now intend to φ. Recalling (in the morning) that yesterday evening I decided to tell my colleague something is not the same as now occurently intending to tell her something. Translating beliefs about my prior decisions into occurrent intentions requires that I actively recommit to the intention. I need to make up my mind again and recommit to the action of telling her something. Conceiving of standing intentions in terms of episodic memory of prior decision-making would therefore systematically require agents to reconsider their plans. This would be in conflict with the claim that future-directed intentions are governed by conduct-control and stability.

Conceiving of standing intentions and their retrieval into working memory in terms of semantic or episodic memory of prior decisions would therefore not allow us to explain how standing intentions can become occurrent again without the need for any re-endorsement or new decision-making.

4.2. Retrospective memory of reasons for action

If standing intentions cannot be explained away as a kind of semantic or episodic memory of prior decision-making, then maybe it can be done in terms of memory of reasons for action. According to this alternative proposal, memory for intention should be understood as the agent's ability to retain and later recall her reason for action. Given our general argument against semantic memory accounts, here we focus only on episodic memory.

Here is one way to articulate this alternative. We can understand practical deliberation as the process by which, given some goal, the agent considers various reasons for actions: reasons for φ-ing, reasons for not φ-ing, reasons for ψ-ing, etc. The agent concludes her deliberation by making up her mind about what she ought to do. In a further step, a rational agent would decide to do what she judges she ought to do and form the corresponding intention. The agent's making a decision is the activity of taking certain considerations "to settle the question of whether so to act" (Hieronymi, 2011) and thereby forming the intention so to act. The intention is thus related to a specific reason to act, namely, the one that decisively speaks in favour of φ-ing (rather than not φ-ing, ψ-ing, etc.). In the case of future-directed deliberation and decision-making, the agent commits her reason for action

to memory. Later, when the time comes, she recalls her reasons. This leads her to form the corresponding occurrent intention.

Consider an example. In the evening, the agent considers reasons for and against telling her colleague about the new coffee machine when meeting her in the morning. She could tell the colleague now by phoning her but it is too late; it would be better to do it in the morning since it is nice to have a few things to say when meeting one another in the hallway, etc. Considering these reasons, she settles on telling the colleague when meeting her in the morning and thereby forms the corresponding intention. She thereby also commits to memory her reasons for telling her in the morning. When seeing the colleague in the hallway, she recalls her concluding reasons – that is, she recalls that it is nice to have something to say, that the colleague might not know about the new coffee machine, that the colleague loves coffee, etc. Consequently, she forms the occurrent intention to tell her about the coffee machine.

In order for this alternative account to work, it is important that recalling one's reasons and forming the corresponding occurrent intention does not involve reconsideration and new decision-making. If recalling one's reasons and re-forming the occurrent intention did involve new decision-making, then we would run into Bratman's puzzle about the rationality of future-directed decision-making. Deliberation about the future consumes time and cognitive resources. If the agent always has to deliberate just before forming an occurrent intention, then it would be a waste of cognitive power to do so before the time comes for performance.

There are reasons for thinking that the present alternative account implies that agents are systematically required to reconsider their decisions. Remembering now (in the morning) the reasons I had in the evening for telling my colleague about the coffee machine in the morning does not automatically lead to the occurrent intention to tell her now. From the fact that I now recall the reasons I considered decisive in the evening, it does not follow that I consider them to be decisive now (my colleague might look really busy, for instance). Recalling one's operative reasons requires new decision-making in order to form an occurrent intention. I need to assess whether the reasons are also decisive now. Remembering that my colleague loves coffee, that there is a new machine on the third floor, and that she probably does not know about the machine yet does not mean that I automatically consider the issue settled now. Due to the dynamically evolving world and the discrepancies between prediction and reality, I cannot rely on the considerations I judged to decisively favour an action yesterday also favour it today.

But maybe one's memory of reasons is insulated from reconsideration in the way that Bratman argued for the default non-reconsideration of future-directed intention. When I recall in the morning the reasons I had in the evening to tell my colleague in the morning, I do not review the reasons again. I simply accept the judgement that I ought to tell her. This objection downplays the fact that the world is dynamic and rarely evolves exactly as expected. What in the evening appeared to be a decisive reason for φ-ing might not be so in the morning. If decision-making is about deciding to do what one ought to do in a given context and recalling one's reason is related to an expected context, then the agent should reconsider her reasons when she suspects that the present context is not exactly as expected. This means that if we want an agent who can stick to her intention across changes in context and some variation in the reasons for action, we need to separate the evaluation of one's reasons and the formation of one's intention, such that the intention can persist even if some reasons do not.

Summing up, we have considered two retrospective accounts of memory for intentions. The first account is that when an agent recalls her intention, she is recalling that she made a decision and formed an occurrent intention. Recalling this fact leads to the formation of a corresponding occurrent intention. The second account is that when an agent recalls her intention, she is recalling the reasons for action she considered decisive at the time. Both of these alternative proposals fail for the same reasons. First, if the involved kind of memory is semantic memory, it introduces an epistemic openness that is inconsistent with the non-reconsideration feature of memory for intention. Second, if memory for intention is an episodic kind memory of past decisions or reasons, recalling one's intention would often not (perhaps never) be sufficient for the formation of the corresponding occurrent intention. Recalling that I made a decision to φ or recalling that I had this reason for φ-ing is not the same as now intending to φ. To make this last step, a new endorsement or decision is needed. If that is the case, then these alternative proposals are not able to explain how intentions can become occurrent again without the need for any re-endorsement or new decision-making. That is, semantic and episodic memory accounts fail to satisfy the third explanatory desideratum for theories of memory for intention.

5. A computational theory of intention selection

In this section, we sketch a number of computational principles that can explain (1) the fact that recalling one's intention is a selection process, (2) the fact that the retrieval process has a semi-automatic character, and (3) the fact that retrieval of one's intention can happen without any new

decision-making. Together this set of principles form the Computational Theory of Intention Selection (CTIS). Along the way, we will present some empirical results that support the computational principles of intention selection. The aim here is not demonstrate the correctness of the theory. We settle for the more modest goal of showing that our account could explain desiderata (1)-(3). Given our arguments against the semantic and episodic memory accounts and the possible computational account of selection of intentions, our final aim is to establish that we should accept that memory for intention is a special kind of declarative memory.

The ability to retain standing intentions in long-term memory and making them occurrent in working memory is related to the set of capacities studied by cognitive psychology under the description of prospective memory (for a book length review, see McDaniel and Einstein 2007). Note, however, often "prospective memory" is used to name a specific type of laboratory task (the event-based prospective memory task, see below). Psychological models and explanations of prospective memory are consequently models and explanations of data obtained using this event-based paradigm. The problem is that the relation between memory for intention and the prospective memory task is complicated. In this section, we focus on long-term representation and sketch a possible account of the computational mechanism for retrieval of standing intentions from long-term memory. There are reasons for thinking the prospective memory task does not require participants to use long-term memory. Participants might have enough cognitive resources to be able to maintain a task representation in working memory while monitoring for cues (Smith, 2016). Consequently, existing computational models of prospective memory are models of data from the event-based prospective memory task (Boag et al., 2019; Horn et al., 2011; Smith & Bayen, 2005; Strickland et al., 2018) and might not be directly relevant for an account of retrieval of intentions from long-term memory.

Before outlining the basic computational principles of intention selection, that is, the retrieval of standing intentions from long-term memory, we need some account of the representation of intentions in memory. When an agent decides to do something in the future, she encodes her future-directed intention into memory. Here is one possible account of how future-directed intentions are represented in memory. An intention is a complex representation of an action to be performed in a certain situation, at a certain time, with certain kinds of objects, and in certain ways. We can group this representational information into three separate components in terms of the relevance to (a) attentional set, (b) motor execution, and (c) the rational place of the action in a hierarchical goal-structure. These different representational components of an intention contain information about: (a) attention: what objects and features to attend to; (b) motor execution: which sensorimotor programs

will be relevant; and (c) rational place in goal hierarchy: the propositionally specified goal of the action. The information in a component can range from very specific and concrete representations to highly schematic and abstract representations.

Take *the attention component*. The attention component is familiar from theories of prospective memory. Imagine an agent who is a participant in a psychological experiment (for instance, in a study of prospective memory). The participant can be instructed to "press F7, when you see the word "tortoise"". To solve such a task, the attentional component represents the letter and word shape features. Or to take another example, if the agent intends to drink from her glass when she returns to her place at the bar, then features of her glass (such as its shape, location, and colour) will be represented. *The motor execution component* is familiar from action-effect studies and theories of ideomotor control (Prinz, Aschersleben, & Koch, 2009). The basic idea is that if I intend to grasp my glass with my right hand, certain motor schemas and their sensory action-effects will be associated with the action. The intention thus contains information about the sensory effects of the action. According to ideomotor control theories, these sensory effects are often sufficient to activate the action representation (Hommel, 2006). Finally, consider *the propositional goal component*. If I intend to write the shopping list for a party, my action will be associated with propositional information about the goal and its place in a larger goal-hierarchy (writing shopping list in order to arrange a party, etc.). The intention contains information about how it fits with higher-order goals.

An intention is represented in long-term memory in virtue of the information in the components. The attentional, the motor execution, and the propositional goal components contain information about how the world should be when the intention is brought to mind and executed. The basic idea behind CTIS is that representational content of the components is matched against the context (that is, against the content in short-term memory). Crudely put, the better the match, the higher the activation of the component (this basic idea will be qualified below). Recall that according to our processing model of memory, the only difference between a standing and an occurrent intention is their level of activation. By activation, we mean that the components are held activated by synaptic reverberating connections (Hebb, 1949/2005; Wang, 2001) between the components and between the components and prefrontal areas of the cortex (Duncan, 2001). An intention becomes occurrent when one of its components becomes ready for use in working memory related processing (reasoning, planning, decision-making, and action control).

Given this sketch of the representational components of intention, we can now state four plausible computational principles that could govern and explain retrieval of intentions from long-

term memory. First, the representations of the components are matched against information from the context as represented in short-term memory stores. This matching process is taking place constantly and should not be confused with an explicit form of resource demanding monitoring for cues. If there is a sufficient match between a component and the context, this will serve as evidence for retrieval. Second, a standing intention becomes occurrent in working-memory when one of its components becomes more active than any component of other intentions. Third, each component comes with a certain prior probability of being activated in a given context. Call this prior probability the bias of a component. If the bias for the attention component is high, it might become active in a given context, even if the match between the information of the attentional component and the context is low. Fourth, not all intentions are equally important to the agent. Some intentions might be so important that the agent keeps thinking of her intention with increasing frequency throughout the day – even if the information in the components find almost no match with the context. Together these principles can explain the semi-automatic selection of an intention from a plurality of standing intentions represented in long-term memory.

Given the account of the components of an intention and the four computational principles, we can describe the selection and retrieval of an intention from long-term memory as depending on three general factors:

1. The *match* between the context as represented in short-term memory and the representations of the components of the intention,
2. The *importance* of the intention relative to other intentions, and
3. The *bias* ascribed to each component.

We can think of Factor 1 as the evidential aspect and Factor 2 as the motivational aspect of an intention (see Brand, 1984). Factor 3 is the probability of activation independent of context match and importance.

Factor 1 is a matching of the representation of a component against the representations currently active in short-term memory. The matching of the attentional component is well-described in research on prospective memory. Take a standard prospective memory (PM) task. In the event-based PM laboratory paradigm, participants are typically instructed to complete a forced choice ongoing task (e.g., lexical decisions; press “F” for word, and “J” for non-word). At the outset of the ongoing task, some participants are instructed to remember to perform a third alternative response (e.g., press the

"F7" key) if they are presented with a PM target event (e.g., a particular word such as *elephant* or a particular syllable, *tra*). If the participant is preparing to react to specific cues for action (in contrast to abstract cues, Ellis & Milne, 1996) or to cues which are similar to cues for an ongoing task (so-called focal cues, in contrast to non-focal cues, Scullin et al., 2010), it increases the probability of remembering to perform the PM task. Furthermore, if the cue-action representation is specified by a process of mental imagery, it will increase the probability of remembering the intention when presented with the cue (Spreng, Madore, & Schachter, 2018). We can explain these effects in terms of how these manipulations facilitate the matching of the representations of the attentional component with the content of perceptual short-term memory. The more specified and focal the representation of the action cue is, the more likely the component is to become active.

Factor 2 is the relative importance of an intention. A match alone is not sufficient to drive the selection and the recalling of the intention. The environment might match many different standing intentions equally well, yet only one is selected – the most important one. In psychological experiments, importance is often operationalized in terms of monetary or social rewards or punishments (Walter & Meier, 2014). For instance, Cook and colleagues showed that situations where the cue representations remain the same between various reward conditions, the probability of remembering the PM task is a function of reward type (the higher the reward, the higher the probability of remembering, Cook, Rummel, & Dummel, 2015). Furthermore, in some situations, there might be a strong match between the environment and the component of an intention, and yet some other intention ends up being selected because it is much more important to the agent. Or imagine a case, where the agent's motivations have changed drastically and executing the intention is no longer relevant to the agent. Even if there is a match, the intention might not be recalled because the importance of the intention is now low. This is the case when participants have been instructed that the PM task is no longer relevant (Scullin, Einstein, & McDaniel, 2009; Scullin & Bugg, 2013).

Factor 3 is a bias ascribed to each of the components describing the probability of activating the specific components independently of the particular intention being processed. Thus, the bias is the probability that the component is active independently of its match with context and the importance of the intention. Agents have some voluntary control over the bias factor for motor execution. Agents can to some extent deliberately turn up or down this bias by either setting the system up for immediate action upon recalling the intention (high bias for the motor component) or making sure that no immediate action will ensue (low bias for the motor component). Take the case of an expert hunter in the wild to shoot pheasants. She might set up her system for fast reaction to the

colour flickering of a certain size at a certain location, so that she can react even before having had time to properly identify consciously the thing as a pheasant (see Verbruggen, McLaren, & Chambers, 2014, on "proactive control"). If the bias for the motor component is set high, the selected intention will lead to immediate action if the operation of the attentional component results in encoding of the appropriate object(s) matching the schema in the motor component.

The agent can voluntarily influence the probability of retrieving the intention not only by adjusting the bias for the motor execution component but also by having some influence on the information associated with an intention. She can clarify and articulate the representations in the components when forming or rehearsing the intention thereby facilitate the matching at retrieval. The agent can do this, for instance, by rehearsing the intention when it is formed and encoded into long-term memory. Imaginative rehearsal of the situations of retrieval and execution is a way to deliberately elaborate or specify the representations of the attention and motor execution components. This would facilitate the matching with the context by hooking the components onto well specified cues. This could raise the probability of recalling the intention when later encountering the cues (Sculling et al., 2017; Spreng et al., 2018).

A standing intention in long-term memory is the disposition of the intention to become occurrent given these three factors (matching, importance, and bias). Formally, these factors are multiplied with each other such that a zero or very small value of one can veto the selection of the intention. For instance, importance might be set low when the agent is informed that it is too late to perform the action. These factors will determine the probability of a component becoming active in working memory in a race against the components of all other standing intentions in long-term memory (they have their values determined in a similar way). Metaphorically speaking, we can think of the biased competition for selection between standing intentions in long-term memory as a horse race where the power of each "intention-horse" is determined by the three factors.[9]

6. Meeting the explanatory desiderata

CTIS is able to satisfy the desiderata for explanations of memory for intentions. First, CTIS provides a computational explanation of how a standing intention is selected from a multitude in long-term memory. It explains the retrieval process as a biased competition between a multitude of potentially context relevant intentions in long-term memory, where one standing intention is selected and becomes occurrent in working memory. When a component wins the biased activation competition,

[9] For related models, see Bundesen, 1990; Desimone & Duncan, 1995; Logan & Gordon, 2001.

it becomes active in working memory together with the other associated components. This makes the intention occurrent as a whole of several components. CTIS's description of long-term representations in terms of components and the account of activation in terms of matching, importance, and bias make sense of the fact that multiple intentions are represented in long-term memory at a low cognitive cost.

Second, CTIS provides us with an explanation of the semi-automatic character of retrieval of intentions from long-term memory. On the one hand, retrieval is not a conscious, deliberate activity but is simply the way in which the three factors regulate the activity level of intentions in memory. When a component becomes active, the intention becomes occurrent and available for rational thinking and action. On the other hand, the agent has some control over the process of retrieval. It is well-known from the literature on prospective memory that agents can influence the retrieval process in various ways, either by strengthening the intention at encoding by imaginative rehearsal (Sculling et al., 2017) or by increasing the likelihood of retrieval by creating external reminders (post-it notes, etc., Gilbert, 2015). These aspects of memory are explained by CTIS in terms of the bias and representational content. According to CTIS, an agent has some control over the bias ascribed to the motor execution component of her intention, and the agent has some control over the specificity and vividness of the information of the components. In this way, the agent has some control over the probability of retrieving the intention and the cognitive role the intention can play if it becomes occurrent.

Third, CTIS provides us with a promising framework for explaining how retrieval of intentions from long-term memory can make intentions occurrent without any need for re-endorsement or new decision-making. Admittedly, this aspect is more like a promising potential than a fully worked out account. The importance (Factor 2, the relative importance of an intention) and the bias (Factor 3, the background probability that a given component will become active) provide us with an explanation of how motivational factors are involved in the selection of standing intentions. However, it is not clear how importance and bias could also explain the fact that when an intention becomes occurrent, it becomes active with the motivational force of an intention. Other features of CTIS are important to the explanation of this aspect of intentions. According to CTIS, intentions often have access to the motor system (depending on the bias and information of the motor execution component) and to rational planning (the propositional goal component). These aspects relate to the "conduct-control" and "rational role" features of intentions highlighted by Bratman (1987). According to Bratman, they

are what is needed in an explanation of the special action-readiness and commitment that sets intentions apart from desires and beliefs.

The force of CTIS in the present context is that it shows us how a non-semantic and non-episodic kind of declarative memory for intentions could be possible. To the extent that philosophers are accustomed to think that semantic and episodic kinds of memory are the only possible kinds of declarative memory (see for instance the taxonomies of declarative memory in De Brigard, 2017, Michaelian, 2016, Werning & Cheng, 2017), this should be interesting on its own. Taken together with the problems that the semantic and episodic memory accounts have in explaining our ability to retrieve intentions without new endorsement and decision-making, we have ground for taking seriously the possibility that memory for intention is a special kind of declarative memory. Following the taxonomic principles outline by Michaelian (2016, Ch. 2), we have grounds for making distinctions between different kinds of declarative memory if there are systematic computational, algorithmic, and implementational differences. In this paper, we have focused on the distinctive computational function played by memory for intention in temporally extended agency. A function that semantic and episodic kinds of memory seem unable to satisfy.

7. Concluding remarks

Assuming a Bratman inspired causal theory of intentions, the distinction between standing and occurrent states, and a specific type of processing model of memory, we argued that a satisfactory account of agents' ability to retain and later recall their future-directed intentions should be able to explain: (1) How the retrieval mechanism solves the problem of selecting an appropriate intention from the multitude of potentially relevant standing intentions; (2) the often semi-automatic character of the retrieval; (3) how standing intentions can become occurrent again without the need for any re-endorsement or new decision-making. We argued that semantic and episodic memory accounts of memory for intentions are unable to satisfy (3). We then presented a computational theory of memory for intentions (CTIS) according to which an intention consists of three representational components. Recalling one's intention is regulated by the way in which the components of all one's intentions are racing towards activation governed by the three factors (matching, importance, and bias).

The ultimate test of CTIS is empirical. Here we attempted to demonstrate the possibility and plausibility of the theory by showing how it fits a number of well-established empirical results from the literature prospective memory and by arguing that it is able to meet the three desiderata. We hope

that this suffices to show that one important form of memory for intention might be its own kind of declarative memory.

Acknowledgements

We are indebted to Chiara Brozzo, Denis Buehler, John Duncan, Mark Schram Christensen, Sam Gilbert, Marion Goodman, Julia Haas, Alexander Heape, Andrew Heathcote, Ben Henke, Gordon Logan, Krisztina Orban, Hong Yu Wong, Wayne Wu, and two reviewers for the journal for insightful comments and very helpful discussions of the material. The ideas have been presented at workshops and seminars at ANU, Cambridge, Copenhagen, Grenoble, Tübingen, Munich, and Aarhus, and we are grateful for the feedback we received on those occasions.

References

Altmann, E. M., & Trafton, J. G. (2002). Memory for goals: An activation-based model. *Cognitive science*, *26*(1), 39-83.

Alston, W. (1967). Motives and Motivation. In P. Edwards (ed.), *The Encyclopedia of Philosophy*. New York: Macmillan.

Anderson, J. R.(1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Atance, C. M., & O'Neill, D. K. (2001). Episodic future thinking. *Trends in cognitive sciences*, *5*(12), 533-539.

Audi, R. (1994). Dispositional beliefs and dispositions to believe. *Noûs*, *28*(4), 419-434.

Boag, R. J., Strickland, L., Loft, S., & Heathcote, A. (2019). Strategic attention and decision control support prospective memory in a complex dual-task environment. *Cognition*, *191*, 103974.

Boyer, P. (2008). Evolutionary economics of mental time travel? *Trends in cognitive sciences*, *12*(6), 219-224.

Brand, M. (1984). *Intending and Acting: Toward a Naturalized Action Theory*. Cambridge, MA: MIT Press.

Bratman, M. (1987). *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.

Bundesen, C. (1990). A theory of visual attention. *Psychological review*, *97*(4), 523-547.

Cook, G. I., Rummel, J., & Dummel, S. (2015). Toward an understanding of motivational influences on prospective memory using value-added intentions. *Frontiers in human neuroscience*, *9*, 278.

Cowan, N. (1993). Activation, attention, and short-term memory. *Memory & Cognition*, *21*(2), 162-167.

Cowan, N. (2019). Short-term memory based on activated long-term memory: A review in response to Norris (2017). *Psychological bulletin*, *145*(8), 822-847.

De Brigard, F. 2017. Memory and imagination. In *Routledge handbook of philosophy of memory*, ed. S. Bernecker and K. Michaelian, 127–140. Abingdon: Routledge.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual review of neuroscience*, *18*(1), 193-222.

Duncan, J. (2001). An adaptive coding model of neural function in prefrontal cortex. *Nature reviews neuroscience*, *2*(11), 820.

Ellis, J., & Milne, A. (1996). Retrieval cue specificity and the realization of delayed intentions. *The Quarterly Journal of Experimental Psychology Section A*, *49*(4), 862-887.

Fernandez, J. (2006). The intentionality of memory. *Australasian Journal of Philosophy*, *84*(1), 39-57.

Fernández, J. (2014). Memory and immunity to error through misidentification. *Review of Philosophy and Psychology*, *5*(3), 373-390.

Fuster, J. M., & Alexander, G. E. (1971). Neuron activity related to short-term memory. *Science*, *173*(3997), 652-654.

Gilbert, S.J. (2015). Strategic offloading of delayed intentions into the external environment. *Quarterly Journal of Experimental Psychology*, **68**, 971-992.

Goldman, A. I. (1970). *Theory of human action*. Princeton, NJ: Princeton University Press.

Guynn, M. J. (2003). A two-process model of strategic monitoring in event-based prospective memory: Activation/retrieval mode and checking. *International Journal of Psychology*, *38*(4), 245-256.

Hebb, D.O. 2005. *The organization of behavior: A neuropsychological theory*. Mahwah, NJ: Lawrence Erlbaum Associates.

Hieronymi, P. (2011). Reasons for Action. *Proceedings of the Aristotelian Society*, 111(3), 407-427.

Hommel, B. 2006. How we do what we want: A Neuro-cognitive perspective on human action planning. In *Planning in intelligent systems: Aspects, motivations and methods*, ed. W. van Wezel, R. Jorna, and A.

Meystel, 27–56. New York: Wiley.

Hommel, B., & Wiers, R. W. (2017). Towards a unitary approach to human action control. *Trends in cognitive sciences*, *21*(12), 940-949.

Horn, S. S., Bayen, U. J., & Smith, R. E. (2011). What can the diffusion model tell us about prospective memory?. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *65*(1), 69-75.

Klein, S. B. (2013). The temporal orientation of memory: It's time for a change of direction. *Journal of Applied Research in Memory and Cognition*, *2*(4), 222-234.

Kvavilashvili, L., & Mandler, G. (2004). Out of one's mind: A study of involuntary semantic memories. *Cognitive psychology*, *48*(1), 47-94.

Logan, G. D., & Gordon, R. D. (2001). Executive control of visual attention in dual-task situations. *Psychological review*, *108*(2), 393-434.

McDaniel, M. A., & Einstein, G. O. (2000). Strategic and automatic processes in prospective memory retrieval: A multiprocess framework. *Applied Cognitive Psychology*, *14*(7), S127-S144.

McDaniel, M.A., and G.O. Einstein. 2007. *Prospective memory: An overview and synthesis of an emerging field*. Thousand Oaks: Sage Publications.

Mele, A. R. (2003). *Motivation and agency*. Oxford: Oxford University Press.

Mele, A. R. (2007). Persisting intentions. *Noûs*, *41*(4), 735-757.

Michaelian, K. (2016). *Mental time travel: Episodic memory and our knowledge of the personal past*. Cambridge, MA: MIT Press.

Naylor, A. (2011). Remembering-that: Episodic vs. semantic. *Philosophical Psychology*, *24*(3), 317-322.

O'Brien, L. 2003. On Knowing One's Own Actions. In *Agency and Self-Awareness: Issues in Philosophy and Psychology, ed. J. Roessler and N. Eilan, 358–382. Oxford: Oxford University Press.*

O'Brien, L. (2007). *Self-knowing agents*. Oxford: Oxford University Press.

Paul, S. K. (2012). How we know what we intend. *Philosophical Studies*, *161*(2), 327-346.

Prinz, W, Aschersleben, G. & Koch, I. (2009) Cognition and action. In E. Morsella, J. A. Bargh, & P. M. Gollwitzer (Eds.), *The psychology of action, Volume 2: Mechanisms of human action* (pp. 35-71). Oxford: Oxford University Press.

Reese, C. M., & Cherry, K. E. (2002). The effects of age, ability, and memory monitoring on prospective memory task performance. *Aging, Neuropsychology, and Cognition*, *9*(2), 98-113.

Schacter, D. L., & Addis, D. R. (2009). Remembering the past to imagine the future: A cognitive neuroscience perspective. *Military Psychology*, *21*(S1), S108.

Schacter, D. L., and E. Tulving. (1994). What are the memory systems of 1994? In D. L. Schacter and E. Tulving (eds.), *Memory Systems*, 1 – 38. Cambridge, MA: MIT Press.

Scullin, M.K., and J.M. Bugg. 2013. Failing to forget: Prospective memory commission errors can result from spontaneous retrieval and impaired executive control. *Journal of Experimental Psychology. Learning, Memory, and Cognition* 39 (3): 965–971.

Scullin, M. K., Einstein, G. O., & McDaniel, M. A. (2009). Evidence for spontaneous retrieval of suspended but not finished prospective memories. *Memory & Cognition*, *37*(4), 425-433.

Scullin, M. K., Kurinec, C. A., & Nguyen, K. (2017). The effects of implementation intention strategies on prospective memory cue encoding. *Journal of Cognitive Psychology*, *29*(8), 929-938.

Scullin, M. K., McDaniel, M. A., Shelton, J. T., & Lee, J. H. (2010). Focal/nonfocal cue effects in prospective memory: Monitoring difficulty or different retrieval processes?. Journal of Experimental Psychology: Learning, Memory, and Cognition, 36(3), 736.

Shoemaker, S. (1970). Persons and their pasts. *American Philosophical Quarterly*, *7*(4), 269-285.

Smith, R. E., & Bayen, U. J. (2005). The effects of working memory resource availability on prospective memory: A formal modeling approach. *Experimental Psychology*, *52*(4), 243-256.

Smith, R.E. (2016). Prospective memory: A framework for research on metaintentions. In J. Dunlosky and S. Tauber (Eds.), *The Oxford handbook on Metamemory*. New York: Oxford University Press.

Spreng, R. N., Madore, K. P., & Schacter, D. L. (2018). Better imagined: Neural correlates of the episodic simulation boost to prospective memory performance. *Neuropsychologia*, *113*, 22-28.

Squire, L. R. (2004). Memory systems of the brain: a brief history and current perspective. *Neurobiology of learning and memory*, *82*(3), 171-177.

Strickland, L., Loft, S., Remington, R. W., & Heathcote, A. (2018). Racing to remember: A theory of decision control in event-based prospective memory. *Psychological Review*, *125*(6), 851-887.

Tulving, E. 2002. Episodic memory: From mind to brain. *Annual Review of Psychology 53* (1): 1–25.

Verbruggen, F., McLaren, I. P., & Chambers, C. D. (2014). Banishing the control homunculi in studies of action control and behavior change. *Perspectives on Psychological Science*, *9*(5), 497-524.

Walter, S., & Meier, B. (2014). How important is importance for prospective memory? A review. *Frontiers in psychology*, *5*, 657.

Wang, X. J. (2001). Synaptic reverberation underlying mnemonic persistent activity. *Trends in neurosciences*, *24*(8), 455-463.

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