A new cognitive model of long-term memory for intentions

Grünbaum, Thor; Oren, Franziska; Kyllingsbæk, Søren

Published in:
Cognition

DOI:
https://doi.org/10.1016/j.cognition.2021.104817

Publication date:
2021

Document version
Early version, also known as pre-print

Citation for published version (APA):
Grünbaum, T., Oren, F., & Kyllingsbæk, S. (2021). A new cognitive model of long-term memory for intentions. *Cognition*, *215*, [104817]. https://doi.org/10.1016/j.cognition.2021.104817

Preprint. Accepted 2021 in *Cognition*.
Please quote the published version: https://doi.org/10.1016/j.cognition.2021.104817

A New Cognitive Model of Long-Term Memory for Intentions

By Thor Grünbaum[1,3], Franziska Oren[2,3], & Søren Kyllingsbæk[2,3]

1 Section of Philosophy
University of Copenhagen
Karen Blixens Plads 8
DK-2300 Copenhagen S.
Denmark

2 Department of Psychology
University of Copenhagen
Øster Farimagsgade 2A,
DK-1353 Copenhagen K
Denmark

3 CoInAct Research Group
University of Copenhagen
Øster Farimagsgade 2A,
DK-1353 Copenhagen K
Denmark

Corresponding author: Thor Grünbaum, tgr@hum.ku.dk

Abstract

In this paper, we propose a new mathematical model of retrieval of intentions from long-term memory. We model retrieval as a stochastic race between a plurality of potentially relevant intentions stored in long-term memory. Psychological theories are dominated by two opposing conceptions of the role of memory in temporally extended agency – as when a person has to remember to make a phone call in the afternoon because, in the morning, she promised she would do so. According to the Working Memory conception, remembering to make the phone call is explained in terms of the construction and maintenance of intentions in working-memory. According to the Long-Term Memory conception, we should explain the episode in terms of an ability to store intentions in long-term memory. The two conceptions predict different processing profiles. The aim of this paper is to present a new mathematical model of the type of memory mechanism that could realise the long-term memory representations of intentions necessary for the Long-Term Memory conception. We present and illustrate the formal model and propose a new type of experimental paradigm that could allow us to test which of the two conceptions provides the best explanation of the role of memory in temporally extended agency.

Keywords:

Intention; Mathematical cognitive model; Long-term memory; Task switching; Prospective memory; Task-set; Temporally extended agency

Preprint. Accepted 2021 in *Cognition*.
Please quote the published version: https://doi.org/10.1016/j.cognition.2021.104817

1. Introduction

In the morning, someone asks you to make a phone call after lunch. You accept, so, in the morning, you decide to make the phone call after lunch. After lunch, you have to remember your intention and make the call.

Psychological theories are dominated by two opposing conceptions of the role that memory can play in this kind of temporally extended agency. We call them the Working Memory (WM) conception and the Long-Term Memory (LTM) conception. According to the WM conception, decision-making and intentions are always formed, maintained, and implemented within working memory. After lunch, you episodically remember that you accepted to make the phone call and you decide to do it now and consequently implement the intention in action. Here long-term memory can only play the role of episodic memory of the event of accepting and what you accepted to do, thus providing you with the ground for making a new decision to make the phone call now.

According to the LTM conception, it is possible for the agent to set herself up for action in the future in such a way that no new decision-making will be necessary. The formation and implementation of intentions might require working memory, but once formed it is possible for the agent to represent intentions in long-term memory. After lunch, you remember your intention and make the phone call. Here long-term memory can play a role in addition to the episodic role; it can store a representation of the intention. When remembering your intention, the standing intention in long-term memory simply becomes active in working memory. No new decision-making needs to be involved and no new intention needs to be constructed. Despite the intuitive plausibility of the LTM conception (Gollwitzer, 1999; Grünbaum & Kyllingsbæk, 2020; McDaniel & Einstein, 2007; Verbruggen, McLaren, & Chambers, 2014), the WM conception remains a dominant picture in cognitive psychology (see, for instance, Grange & Houghton, 2014a, 2014b; Monsell, 2003). Some of the practical and theoretical reasons for this dominance become clear in our discussion task-switching and prospective memory (Section 5).

The aim of this paper is to present a new mathematical model of the type of memory mechanism that could realise the long-term memory of intentions necessary for the LTM conception. We outline formal models corresponding to the two conceptions, derive their conflicting quantitative predictions, and propose a new type of experimental paradigm that could allow us to test which of the two conceptions provides the best explanation of the role of memory in temporally extended agency. Rather than presenting data from the new paradigm, we will in this paper focus on presenting the theory, the formal models, the quantitative predictions, and sketch one possible type of paradigm that would allow critical tests of the predictions.

The paper proceeds as follows. In Section 2, we clarify what we mean by *intention* and relate the notion of intention to the notion of task-set. Section 3 outlines two general features that characterise the long-term representations of task-sets according to the LTM conception: (1) task-sets can be represented in long-term memory without any significant cost to ongoing behaviour, and (2) retrieval of a task-set from long-term memory is a selection from a plurality of potentially relevant task-sets. In Section 4, we formulate a new type of formal model of the selection of intentions from long-term memory (a Model of Intention Selection, MIS) that satisfies the LTM conception and explain the two general features. The formal model framework also enables us to define a model (a WM Conception Model, WM-CM) that would satisfies the WM conception. We use the two competing models to make quantitative predictions. In Section 5, we argue that standard task-switching paradigm and the prospective memory paradigm are unable to test the predictions of the two formal models against each other (MIS vs WM-CM). Section 6 outlines a new type of experimental paradigm that would enable us to test which of the two models is most likely to be correct.

2. Intentions and task-sets

In contrast to automatic and stimulus driven behaviour, flexible and intelligent behaviour is guided by internal representations of goals. In the philosophical literature (Mele, 2009a) and the psychological research on prospective memory (Kvavilashvili & Ellis, 1996), these internal representations are often called *intentions*; whereas the psychological research on cognitive control prefers the notion of a *task-set* (Gibson, 1941). Despite investigating a common phenomenon, different experimental and theoretical traditions use different conceptualizations. To avoid confusion and to embrace a wide theoretical scope, in this section, we clarify how we use the notion of intention and how the notion is related to the notion of task-set. We then reformulate the two competing conceptions of the role of memory in temporally extended agency in terms of task-sets.

Intentions are commitments to perform actions (Bratman, 1987; Mele, 1992, 2009b). A person can form an intention to do something by deciding to do it. If a person decides to make a phone call after lunch, she forms the intention to make the call after lunch. Intentions are thus a special kind of action-plan. These action-plans can represent the action at various levels, ranging from abstract propositional representations to concrete and context-dependent representations of motoric and sensory features (Pacherie, 2008).

Philosophers' notion of intention is related to psychologists' notion of task-set (Vandierendonck, Liefooghe, & Verbruggen, 2010). According to one conception, we should distinguish between representations of tasks and task-sets (Schneider & Logan, 2014). One way to understand this distinction is to say that a representation of a task is a propositional description of a goal, whereas a task-set is a representation of attentional, evidential threshold, and motor parameters needed to achieve the goal (Logan & Gordon, 2001). For instance, the task might be to classify whether a given stimulus is odd or even, and the task-set might then be the set of parameters in the perceptual and motor systems in the brain that needs to be specified to respond odd or even to a given stimulus by pressing a specific key with a finger. According to this conception, the psychologists' notion of a representation of a task corresponds to the philosophers' notion of intention.

We propose an inclusive understanding of the notion of intention as including the notion of task-set. As we use the notion of intention, it equals the representations of *task* plus *task-set*. We will set out our conception of intention more formally in Section 4. Briefly put, we argue that an agent's intention consists of three representational components. One component is the propositional representation of a goal, and the other components are the attentional and motor parameters that need to be specified to perform the action. The propositional goal component corresponds to the representation of the task, whereas the attention component and the motor component correspond to the task-set.

We can now describe the two conceptions of the role of memory in temporally extended agency in terms of task-sets. Recall, according to the WM conception, intention-formation, intention-maintenance, and intention-implementation always take place in working memory. So, according to the WM conception, there is only one task-set at a time. Switching tasks implies that the agent is constructing/reconfiguring the task-set in working memory (Rogers & Monsell, 1995). By contrast, according to the LTM conception, intention-maintenance can be a long-term memory affair. So, according to this conception, many task-sets can be represented in long-term memory, even if only few task-sets can be occurrent in working memory at the same time. Switching tasks could imply that the agent is recalling and activating a task-set stored in long-term memory (Logan & Bundesen, 2003, 2004). Notice that the LTM conception allows that some task situations are best described by maintenance of a task-set in working memory. To tell the LTM and WM conceptions apart, we therefore need to focus on the situations that could involve long-term memory encoding, retainment, and retrieval of task-sets.

3. Capacity limitations, costs, and the selection of task-set

The WM conception of the role of memory in temporally extended agency is an influential conception in cognitive psychology. This becomes clear when looking at the experimental studies of task-switching and prospective memory (see Section 5). Rather than being explicitly stated, the WM conception is often a tacit background assumption in experimental and theoretical literature on cognitive control, which becomes manifest by the fact that studies are restricted to situations with only one task-set at a time (for instance, see Monsell, 2003; Searle, 1983, Ch. 3). Nevertheless, a number of strong theoretical arguments support the LTM conception. In this section, we rehearse some of the theoretical arguments supporting the LTM conception.

Human rationality and decision-making are constrained by various cognitive capacity limitations (Simon, 1957; Gigerenzer & Selten, 2002). Given bounds on human working memory, processing capacity, and processing speed, it is impossible for agents to compute utility functions for all possible actions 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 spent on monitoring and reacting to 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 and place of action (Bratman, 1987).

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 (Broome, 2013, Ch. 10). If intentions lost their power over time, were too easily reconsidered, or always required reconsideration and new decision-making 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 should not bother spending one's time and cognitive resources on future-directed deliberation and decision-making (Bratman, 1987). Basically, if intentions always have to be formed (or task-sets constructed) in the situation of action, one should not waste psychological resources on forming them ahead of time.

Phenomenologically, we do seem to form intentions to act sometimes hours or days in advance of the situation of action. Given the two conceptions, there are three different possible consequences. *If the WM conception is correct*, then either (1) the phenomenological appearance is deceiving and we do not actually form future-directed intentions or (2) we form future-directed intentions but are irrational by wasting psychological resources on making the same decision twice. *If the LTM conception is correct*, then (3) we form future-directed intentions sometimes well in advance of the situation of action and we are able to represent these intentions in long-term memory and retrieve them when the time is right. Adopting the WM conception thus comes with a theoretical cost. We would have to explain either the deceiving phenomenology or our irrationality.

Given the LTM conception and a number of standard background assumptions about human cognition, we can infer a couple of features that long-term memory for intentions should be expected to have:

a) *Low cognitive cost*: Intentions are represented in long-term memory at a low cognitive cost. This follows from the human ability to successfully plan, coordinate, and execute future-directed intentions. This ability would be thwarted or severely limited if intentions could only be retained as occurrent in working memory. It is generally accepted in research on prospective memory that agents can maintain intentions or task instructions occurrent in working memory only at a cost to ongoing performance (Scullin, McDaniel, & Einstein, 2010; Smith, 2003; Smith & Bayen, 2005; Smith et al., 2007).

b) *Multiplicity and selection*: At any given time, multiple standing intentions are represented in long-term memory. This follows from the delay between decision and execution, the low-cost representations in long-term memory, and the for practical purposes unbounded nature of long-term memory. When forming a future-directed intention, an agent does not need to check the storage capacity or wait until she has performed some of the intentions already stored. The agent can ordinarily simply rely on her unbounded capacity for long-term memory for intentions. Consequently, at any given time, a multiplicity of intentions is represented in long-term memory. Retrieval of intentions from long-term memory therefore involves a selection process: some mechanism selects the right intention to become occurrent out of a plurality of standing intentions in long-term memory (Grünbaum & Kyllingsbæk, 2020).

Summing up, the limitations of human cognitive capacities imply that if we form intentions for the future, the intentions should be able to persist in some long-term memory format. To be sure, we might deny that humans make future-directed decisions or insist that they are irrational in doing so. But if we accept that humans are rational and make future-directed decisions, we should expect the outcomes of the decisions (i.e., the intentions) to persist in long-term memory at a low cognitive cost and in a large number.

The LTM conception does not assume that an agent is always using long-term memory to solve a task or that using long-term memory is always the most efficient strategy. Rather, it claims that agents are sometimes using long-term memory. To test the two conceptions, the challenge is therefore to design experiments that encourage and constrain long-term memory strategies for encoding, retaining, and retrieving task-sets. Assuming a task situation where a long-term memory strategy should be prevalent, the WM conception predicts that the cognitive processes implement a task-set construction or reconfiguration, whereas the LTM conception predicts that the cognitive processes could implement a selection process from long-term memory. According to the LTM conception, there is always a large number of task-sets represented in long-term memory, and the job is to select the right task-set for activation. Activating a task-set in long-term memory is therefore a selection process. If we can find evidence for a selection process that is sensitive to the number of potentially relevant task-set representations in long-term memory, we would have a reason for choosing the LTM conception.

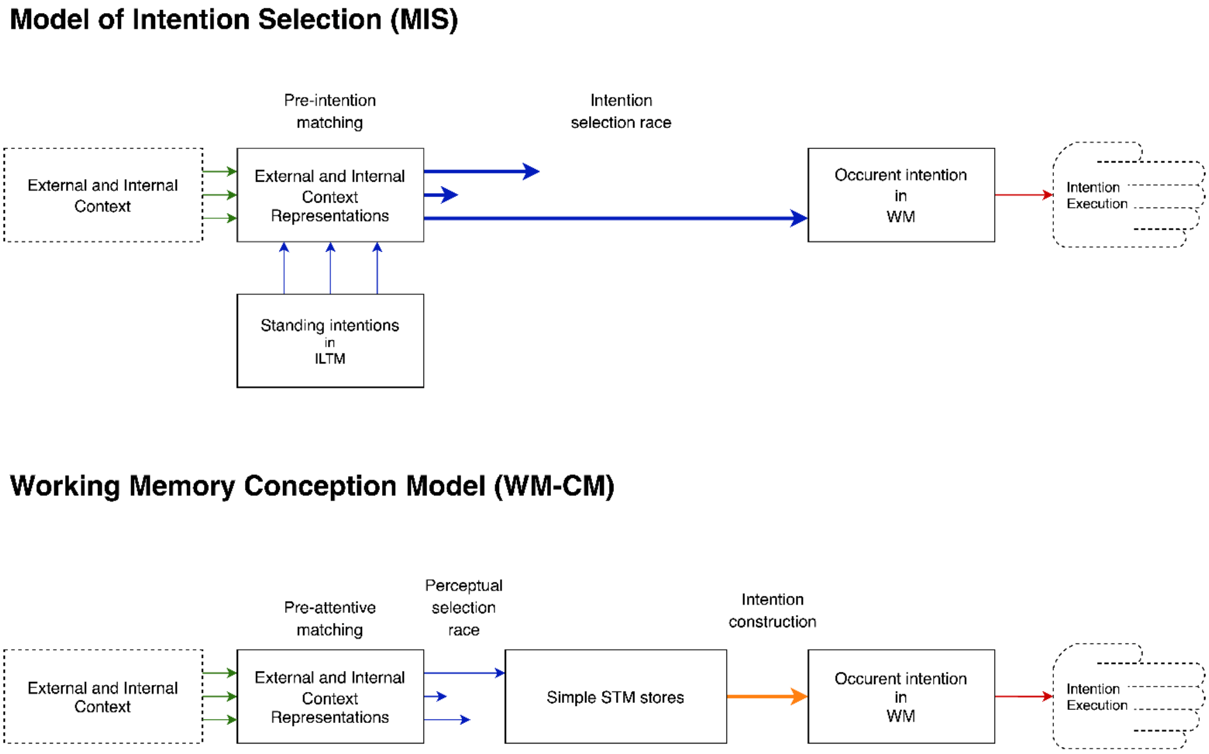

Figure 1. Flow diagrams of the Model of Intention Selection (MIS) and the WM Conception Model (WM-CM). In MIS, the standing intentions in ILTM are first matched against representations of the external and internal context. This triggers a selection race between the possible context component representations where the first intention wins the race by being selected into WM as an occurrent intention. If the attention/decision component is encoded and is matched with the context above threshold and the motor/cognitive component of the occurrent intention is encoded in WM, the intention will be executed as either a covert cognitive operation or an overt motor action. In WM-CM, the selection race is only a perceptual selection race initiated after a pre-attentive perceptual matching process between the external and internal context and representations in long-term memory. The race leads to the selection of a small number of representations in different modality specific short-term memory systems (e.g., Baddeley's working memory model, 2000). Based on this information, the occurrent intention is constructed in working memory and subsequently executed.

4. New model of intention selection: MIS

In this section, we sketch a new mathematical model of the selection of intentions represented in long-term memory. We call this account the Model of Intention Selection (MIS[1]). MIS models the time course of intention retrieval and can explain (a) the representation of standing

[1] In Grünbaum & Kyllingsbæk (2020), we called it the Computational Theory of Intention Selection (CTIS).

intentions in long-term memory at a low ongoing cognitive cost, (b) the representation of multiple standing intentions and retrieval as a selection process, and (c) the role of motivation in the retrieval of standing intentions into working memory. To signify that the intentions are stored in a special type of long-term memory, we refer to this store as the *Intention Long-Term Memory* (ILTM), which is functionally different from other kinds of declarative memory, e.g., episodic and semantic memory (Grünbaum & Kyllingsbæk, 2020). In Figure 1, we have illustrated both MIS and the WM-CM in two separate flow diagrams (we introduce WM-CM in Section 4.3 below).

4.1. MIS: Components, factors, and selection

According to MIS, an intention is a complex representation of an action to be performed in a certain situation, with certain kinds of objects, and in certain ways (for instance, you might intend to wave at a person the moment she enters the room). We call these different information clusters the *components* of the intention. In addition to the components, MIS associates to the intention a *weight* representing the importance of the intention and a *bias* ascribed to each of the components.

According to MIS, an intention consists of three types of components: (a) the attention and decision component; (b) the motor and cognitive execution component; and (c) the propositional goal component. The attention and decision component and the motor and cognitive execution component correspond to the task-set, whereas the propositional goal component corresponds to the propositional representation of the task. The information in a component can range from very specific and concrete representations to highly schematic and abstract representations, e.g., for higher order intentions in the goal hierarchy. A standing intention is represented in ILTM in virtue of the information in its components and the strength of the associations between the components.

The attention/decision component contains information that specifies what objects and features to attend to, as well as thresholds for gathering of perceptual evidence. The content of the attention component will be matched against perceptual representations of objects and features in the perceptual context. If you intend to wave at your friend the moment she enters the room, your attentional system will be tuned the facial features of your friend which will be matched with a representation in the component of the intention. Furthermore, the agent will set a threshold for when she has gathered enough information. For instance, if it is a high stakes case and visually selecting the wrong object could have bad consequences (waving at the wrong person in a crowded room), the agent will continue to gather visual evidence for a longer time to minimize misidentification (see Blurton, Kyllingsbæk, Nielsen, & Bundesen, 2020; Christensen, Markussen, Bundesen, & Kyllingsbæk, 2018; Kyllingsbæk, Markussen, & Bundesen, 2012).

The motor/cognitive execution component specifies which sensorimotor programs and cognitive operations will be relevant to execution. If I intend to wave 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 (Hommel, 2006; Prinz, Aschersleben, & Koch, 2009). For more cognitive intentions, the execution component can be a covert cognitive operation (such as mental imagery, mental math, or the construction, manipulation, and storage of intentions) rather than an overt action.

Finally, *the propositional goal component* contains information specifying propositional representation of the rational place of the task in a goal hierarchy. If you intend to wave at a person in order to attract her attention in order to give her a message, your action will be associated with propositional information about the goal and its place in a larger goal-hierarchy. This hierarchical aspect of intentions is more familiar from classical AI theories (Newell, 1990) and cybernetic theories of motivation and action-control (Austin & Vancouver, 1996) than theories of cognitive control and prospective memory.

The attention/decision, motor/cognitive, and propositional goal components contain information about how the context should be when the standing intention is retrieved. The representations of the components are matched against information from the context. The context can be both internal, e.g., in the form of other intentions, thoughts, and feelings, and external, e.g., in the form of the "gist" of the visual scene or the currently active representations of the context (see Figure 1; Kadohisa et al., 2020; Olivia, 2005; Vidal et al., 2005). According to MIS, this matching process is continuously happening and should not be confused with an explicit form of resource demanding monitoring for cues.

The matching of the attention/decision component is described in the literature on prospective memory (see Section 5 for details about the experimental task). If the prospective memory (PM) task representation involves specific cues for action (Ellis & Milne, 1996) or cues with substantial processing overlap with cues for the ongoing task (focal cues, in contrast to abstract 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 correctly when presented with the cue (Spreng, Madore, & Schachter, 2018, see also Gollwitzer, 1999). We explain these effects in terms of how the manipulations facilitate the matching of the representations of the attention/decision component with the representations of the context. The more specified and focal the representation of the task cue is, the more likely the component is to become active upon encountering the cue. The experimental literature has predominantly focused on visual cues but the matching of the content of the motor/cognitive executive component with perceptual features of the context has been studied by research on motor control in the ideo-motor tradition (Prinz, Aschersleben, & Koch, 2009). The propositional goal component has been studied in classical AI, where goal hierarchies are sometimes described as a matching of a goal with other goal representations (Sacerdoti, 1974).

Thus, the basic idea of MIS is that a standing intention in ILTM becomes occurrent in working memory when its components match the context and are selected. However, the match between a component and the context is just one factor determining the likelihood that the intention will be selected – two other factors are also relevant: importance of the intention and biases related to the components. Given this account of the components of an intention, we can describe the selection and retrieval of an intention from ILTM as depending on three general factors:

1. The *match* between representations of the context and the representations in ILTM of the components of the standing intention;
2. The *importance* of the intention relative to other intentions (formalised as a weight); and

3. The *bias* ascribed to each component instance.

Mathematically, the three factors are independent of each other. We can think of Factor 1 as the evidential aspect and Factor 2 as the motivational aspect of an intention. Factor 3 is simply the degree of importance of a component instance independent of context match and importance.

Factor 2 is the relative importance of an intention. A match alone is not sufficient to drive the selection and retrieval of the intention. The environment might match many different standing intentions equally well, yet only one is selected – likely the most important one. For instance, Cook and colleagues showed that situations where the cue representations remain the same between various reward conditions and there is no significant processing cost to ongoing behaviour, the probability of remembering the PM task is influenced by reward type (Cook, Rummel, & Dummel, 2015). 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 is unlikely to 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). According to MIS, the importance of an intention is influencing the selection proportionally to the sum of all other intentions rather than as a weight with an absolute value (see below, Equation 1). This in turn limits the total processing capacity of the intentions (see below, Equation 2, and Kyllingsbæk, 2016). Thus, if many intentions have a high weight, the time to process each of the individual intentions will increase with the number of intentions (see Figure 2 and Section 4.2).

Factor 3 is a bias ascribed to each of the components describing the likelihood of activating the specific component independently of the particular intention being processed. Whereas the Factor 2 ascribes a weight to the whole intention, we can think of the bias as a measure of how important the component instantiation is to the agent. If the bias for motor/cognitive component is set high, the selected intention will lead to immediate action if the operation of the attention/decision component results in encoding of the appropriate object(s) matching the schema in the encoded motor/cognitive component. This gives the model a way to explain so-called commission errors (Bugg & Streeper, 2019). In situations where the PM instructions are no longer relevant, a match between the attention/decision component and a cue and a high bias for motor/cognitive execution components could lead to PM performance.

Agents have some voluntary control over the bias factor for the component for motor/cognitive 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/cognitive components) or making sure that no immediate action will ensue (low bias for the motor/cognitive components). For instance, an expert hunter 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"). One possibility is to think of the difference between a low and a high bias for the motor/cognitive execution components in terms of a difference between declarative and procedural representations of task instructions (Brass et al., 2017). Turning up the bias for the motor/cognitive execution component raises the probability that the intention becomes active in working memory with its procedural representation of the action. We remain agnostic about whether agents have any deliberate control over the bias for the two other types of components.

Summing up, a standing intention in long-term memory is disposed to become occurrent in working memory given the three factors (matching, importance, and bias). Formally, these factors are multiplied with each other such that a value close to zero of any one of the three can veto the selection of the intention (see Equation 1 below). The three factors determine the rate and thus the probability of the encoding of a component into working memory in a race against all other standing intentions in ILTM. Even if, in principle, all standing intentions stored in ILTM are constantly being matched and racing against each other, most intentions will in a given situation have little contextual support and an importance close to zero. In practice, only a small subset will effectively enter the race for encoding into working memory. The notion of a selection race will be formally explained in the next section.

4.2. MIS: Mathematical formalisation

Formally, the three factors (matching, importance, and bias) are represented by a single unified *rate equation* specifying the rate of processing, $\lambda(\chi, k)$, of component k of intention χ:

$$\lambda(\chi, k) = \psi(\chi, k)\mathrm{B}_k \frac{\omega_\chi}{\sum_{z \in I} \omega_z}, \qquad (1)$$

where $\psi(\chi, k)$, the matching parameter, is the contextual support in the environment for intention χ having component $k \in K$, and K is the set of the three components $\{k_{\mathrm{AD}}, k_{\mathrm{MC}}, k_{\mathrm{P}}\}$, i.e. (AD = attention/decision, MC = motor/cognitive, and P = propositional). As described above, the $\psi(\chi, k)$ values are computed by matching representations of the component of the standing intention in ILTM with representations of the present situational context.

The importance factor, $\omega_\chi / \sum \omega$, represents the relative weight of intention χ compared to the set I of all standing intentions. Thus, the individual weights for each intention, ω_χ, enter the equation relatively to the sum of the weights of all intentions in ILTM, $\sum \omega$. This effectively limits the sum of all $\lambda(\chi, k)$ values, which corresponds to the total processing capacity, C_I, of all the standing intentions participating in the selection race:

$$C_I = \sum_{\chi \in I} \sum_{k \in K} \lambda(\chi, k), \qquad (2)$$

Where again I is the set of all intentions and K is the set of all three components. The limited cognitive capacity is formally implemented in this way (for related models, see Kyllingsbæk, 2016; Logan & Gordon, 2001).

Finally, the bias factor is represented by parameters, B_k, the bias towards activating component k independently of the matching and the importance of intention χ.

Preprint. Accepted 2021 in *Cognition*.
Please quote the published version: https://doi.org/10.1016/j.cognition.2021.104817

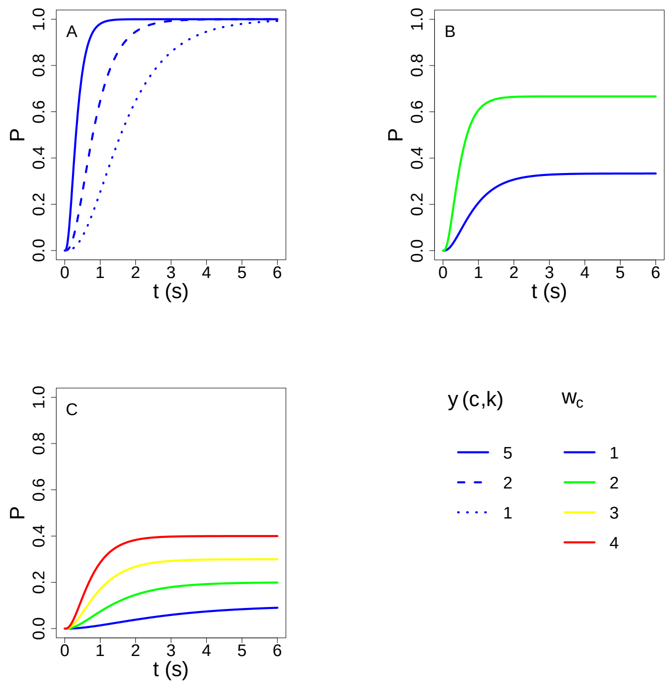

Figure 2. Examples of the encoding time of intentions being selected into working memory. Each curve represents the probability, *P*, of encoding the intention first, i.e., its three components, as a function of time, *t*, measured in seconds (see Equation 3). In Panel A, *blue solid*, *dashed*, and *dotted* lines indicate processing rates of five, two, and one component per second for a single intention racing alone to be encoded into working memory. In Panel B, *blue* and *green* curves indicate competitive processing of two intention components with weights of one and two, respectively. In Panel C, *blue*, *green*, *yellow*, and *red* curves indicate processing of intention components with weights equal to one, two, three, or four, respectively.

Figure 2 illustrates examples of the time course of the selection race between intentions into working memory. An intention wins the selection race when one of its components is encoded as the first one. The rest of the components of the winning intention may then be encoded into WM. Any component from any of the other intentions that finishes processing are lost.

In Panel A, the effect of differences in processing rates are illustrated when a single dominating intention is encoded into working memory. By dominating, we mean that the weight of this intention is so much higher than the weight of any other intention that the ratio of weights in Equation 1, $\omega_\chi/\sum\omega$, equals one. In this case, the rate of processing, $\lambda(\chi, k)$ of component k of intention χ, is only determined by the matches to the context, $\psi(\chi, k)$, and the bias values, B_k, across the set of all components K. If we for simplicity assume that all bias values are set to one, then the summed rate of processing of the components of intention χ equals $\lambda(\chi) = \sum_K \psi(\chi, k)$. The three curves show how the rate of processing given by the degree of contextual match is influencing performance. The three curves represent processing rates measured as the number of intention components processed per second, for values of five, two, and one. The higher the rate of processing, the faster the increase in the probability of the intention being selected and encoded, i.e., the difference in steepness of the three curves.

Panels B and C illustrate the race between two and four intentions, respectively. The intentions have weights of one, two, three, and four. The corresponding curves are coloured in *blue, green, yellow, and red,* respectively. Here we assume that the processing capacity is five intention components per second. In Panel B, two intentions with weights of one and two, respectively, are racing against each other to be encoded into working memory. The rate of processing of each of the three components for the two intentions are $\lambda(\chi, k) = 5 \cdot 1 \cdot 1/(1+2) = 5 \cdot \frac{1}{3}$ for the intention with a weight of one and $\lambda(\chi, k) = 5 \cdot 1 \cdot 2/(1+2) = 5 \cdot \frac{2}{3}$ for the intention with a weight of two (see Equation 1).

Assuming for simplicity that processing is independent and exponentially distributed with rates equal to the rate of processing defined in Equation 1, the probability distribution, $P(t)$, of all three components of intention χ have finished processing at time t is

$$P(t; x) = \prod_{k \in K} 1 - \exp[-\lambda(\chi, k)t], \tag{3}$$

Where again K is, the set of the three types of components $\{k_{\text{AD}}, k_{\text{MC}}, k_{\text{P}}\}$, where AD = attention/decision, MC = motor/cognitive, and P = propositional. For further details of the derivation of the time course of selection between several intentions, please see the Appendix.

4.3 An alternative model: WM-CM

An opposing alternative to our MIS is the *WM Conception Model* (WM-CM) described in general terms in Sections 1 and 2 above. The WM-CM states that intentions and task-sets are always constructed, maintained, and implemented within working memory following a perceptual selection race (see Figure 1).

How may we compare and evaluate MIS and WM-CM? Conveniently, we can transform MIS without difficulty into a formal version of WM-CM, corresponding to the WM conception of the role of memory in temporally extended agency. This enables us to make direct mathematical and quantitative comparisons between the two alternative models.

For WM-CM, we formulate mathematically the construction of an intention in working memory by defining the rate of the construction process. We obtain the rate equation of WM-CM by altering the rate equation of MIS (Equation 1) to include an absolute weight rather than a relative weight of intention χ:

$$\lambda(\chi, k) = \psi(\chi, k) \mathrm{B}_k \omega_\chi, \tag{4}$$

where again, $\psi(\chi, k)$, the matching parameters consist of the contextual support in the environment for intention χ in relation to component k. As described above, the $\psi(\chi, k)$ values are computed based on the initial perceptual pre-attentive processing leading to a representation of the context in simple short-term memory systems (e.g., phonological loop and visuo-spatial sketch pad; Baddeley, 2000) (see Figure 1). The bias factor is again represented by parameters, B_k, the bias towards activating component k. The single absolute weight, ω_χ, in Equation 4 implies that one and only one intention is constructed in working memory where the speed of

Preprint. Accepted 2021 in *Cognition*.
Please quote the published version: https://doi.org/10.1016/j.cognition.2021.104817

the construction can depend on motivational factors represented by ω_χ. There is thus no competition between intentions, and intentions are *not* selected from a plurality in long-term memory.

Due to the absence of a relative importance factor in the rate equation of WM-CM, the processing capacity of the construction of the intention in working memory is given by again the sum of the $\lambda(\chi, k)$ values, corresponding to the total processing capacity, C_χ, but now summed only across the set of components of the intention:

$$C_\chi = \sum_{k \in K} \lambda(\chi, k), \tag{5}$$

where K is the set of all components.

Looking again at Figure 2, the WM-CM only predicts data patterns presented in Panel A, now representing the *construction* of the intention in working memory rather than the capacity limited *selection* of an intention from ILTM into working memory. Assuming again for simplicity that processing is exponentially distributed for the construction process also, the probability distribution of the intention being constructed in working memory is again given by Equation 3 (see also the Appendix).

In relation to WM-CM, Panel A illustrates how the rate of construction of the single intention is related to the degree of contextual match given by the different values of $\psi(\chi, k)$, in this case of a total processing capacity, C_χ, of five, two, and one component per second. Again, the higher the rate of processing, the faster the increase in the probability of the intention being constructed in working memory, i.e., the difference in steepness of the three curves.

Notably, the patterns of behaviour exemplified in Panels B and C for the prediction by MIS are *not* readily predicted by the WM-CM. That is, if the behavioural data is best described by the curves in Panel B and C, WM-CM will not be able to model the data simply as a function of the intention construction process in working memory. By contrast, MIS would be able to model the data as a function of the selection of an intention from a plurality in ILTM. Thus, Panels A, B, and C describe behaviour that would differentiate the two models both quantitatively as well as qualitatively. We can exploit this when comparing the models to empirical data. In the next two sections, we sketch possible experimental paradigms that might be used to test the predictions of the two models against each other.

5. Types of experimental paradigm

We have specified two competing formal models (WM-CM vs MIS) that each corresponds to a general conception of the role of memory in temporally extended agency (WM conception vs LTM conception). Confirmation and disconfirmation of the model predictions will give us reason to choose either the WM or the LTM conception. Given the formal assumptions of the models, we can use the models to derive quantitative predictions of the behaviour of participants in relevant experimental paradigms. Importantly, since MIS assumes that temporally extended action can be explained and modelled as a function of a selection of an intention from long-term memory, testing WM-CM vs MIS requires that a relevant experimental paradigm enables us to manipulate the number of intentions stored in long-term memory.

Recall the two features of the proposed selection process at the heart of the LTM conception. First, task-sets can be represented in long-term memory at a low cognitive cost to ongoing thought and action. Second, a plurality of task-sets can be represented in long-term memory. Activation of a task-set therefore presupposes a selection of the right task-set from a plurality of task-sets. Only if an experimental paradigm can be used to manipulate the selection of task-sets from a plurality of standing task-sets in long-term memory, can the paradigm be used to test the predictions of WM-CM against those of MIS.

The natural place to look for relevant experimental paradigms and results to test the models is in the task switching literature. Imagine that you are waiting for an email. You decide to look at your emails if the 'received an email'-icon becomes visible. Later, while entirely engaged in preparing slides for an upcoming lecture, you see the icon, and you recall your intention to read your email. You stop writing on your lecture slide and proceed to click on the email icon. You switch from one task to the other.

We have focused on two conceptions of this type of task switching. According to the WM conception, given the cue, the task-set for clicking the email icon is constructed in working memory. Task switching comes down to the process in which the task-set involved in writing lecture slides is reconfigured into a task-set for clicking on the email icon (Rogers and Monsell, 1995; Monsell, 2003). According to the LTM conception, it is possible that you have a task-set for clicking the email icon stored in long-term memory. Given the cue, a process starts that ends by the system selecting the clicking-the-email-icon task-set among a plurality of possible task-sets represented in long-term memory (Logan & Bundesen, 2003, 2004).

Task-switching paradigms are as a general class characterized by giving participants two (or more) tasks and instructing them to switch between them according to some schedule or cue. Here we will distinguish between standard task-switching paradigms and prospective memory (PM) paradigms, where PM paradigms are characterized by the fact that one task is occurring only a few times. In our discussion, we will focus on the following two questions: (1) Do the procedures manipulate task-sets *represented in long-term memory*? (2) Do the procedures manipulate the *selection of a task-set from a plurality* (i.e., more than two) of task-sets in long-term memory? Only if these two conditions are satisfied, can we use an experimental paradigm to test WM-CM against MIS.

5.1 Standard task-switching paradigms

These procedures are standardly used to study the cost (extra time needed) involved when the switching between tasks is compared to task repetitions. Since the various procedures and results have already been reviewed thoroughly in the literature (Grange & Houghton, 2014a; Kiesel et al., 2010; Vandierendonck, Liefooghe, & Verbruggen, 2010), we will be brief.

In the *list procedure* (e.g., Allport et al., 1994; Spector & Biederman, 1976), the time needed to complete a "pure list" (AAAA…, for instance, add 3 to each number) is compared to the time needed to complete an "alternating list" (ABAB…, for instance, alternate between +3 and -5). In the *alternating runs procedure* (Monsell, Sumner, & Waters, 2003; Rogers & Monsell 1995; Sumner & Ahmed, 2006), each task is repeated a number of times before switching to the other task (AAABBBAAA…, for instance, alternating every third trial between odd-even decision and vowel-consonant decision). The cost of task switching is measured by comparing switch trials (AB and BA) with repetition trials (AA and BB). In the *explicit cuing procedure* (Meiran, 1996, 2014; Sudevan & Taylor, 1987), on each trial, a task cue indicates which task is to be performed. The switch cost is measured by comparing tasks where the preceding trial was different with trials where the preceding one was the same in a randomized series of trials (for instance, AABABBABAAA…). Compared to the other procedures, the explicit cuing procedure allows for better control of time parameters in processing and performance of the tasks.

These procedures have produced a wealth of results, theorising, and modelling of cognitive control involved in task switching. For our purposes the main issue is whether the procedures can be used to study the selection of task-sets from a multitude of task-sets in long-term memory. The answer is most likely no. The standard task-switching procedures do not target long-term memory. The procedures invite processing strategies where participants can maintain the two task instructions or task-sets in working memory. On all three procedures, the two tasks are presented equally often. Furthermore, the procedures cannot be used to study selection from a plurality of task-sets in long-term memory. Even if the task that is not currently being performed is assumed to be in long-term memory, the task currently performed is in working memory. Consequently, there could at most be one task-set relevant to the experimental task in long-term memory (not currently in working memory), so no selection among a plurality. In sum, the standard task-switching paradigm cannot be used to test the MIS against WM-CM. In terms of the model curves (Figure 2), MIS and WM-CM both predict a curve corresponding to a single intention (Figure 2A).

5.2 Prospective memory (PM) paradigm

The PM paradigm is characterised by the infrequent occurrence of PM targets. This fact together with a number of task features used to encourage long-term memory processing strategies might enable this paradigm to study the selection of task-sets from a plurality of task-sets represented in long-term memory. As we will see, this conclusion might be false.

In the event-based PM laboratory paradigm (Einstein & McDaniel, 1990), participants are typically instructed to complete a forced choice ongoing task (e.g., lexical decisions; press "F" for word and "J" for nonword). At the outset of the ongoing task, some participants are instructed to remember to perform a third alternative response (the PM task, e.g., press the "F7" key) if they are presented with a PM target (e.g., a particular word). Typically, researchers are interested in participants' performance of the PM task in response to the PM cue as well as comparing participants' performance on the ongoing task with and without the additional PM instruction. This basic paradigm has been modified in many ways (Rummel & McDaniel, 2019).

The PM paradigm is ideal for investigating processing costs with or without PM instructions. The difference between, on the one hand, maintaining an occurrent intention in working memory and, on the other hand, storing and spontaneously retrieving a standing intention from long-term memory is standardly operationalized in terms of a processing cost to ongoing behaviour. If the agent is maintaining the intention in working memory and actively monitoring the context for cues, then we should expect a cost to ongoing behaviour during the maintenance phase (usually in terms of longer response latencies). If the agent has successfully stored the intention in long-term memory and is spontaneously retrieving it on encountering the cue for retrieval, then we should expect little processing costs to ongoing behaviour during the maintenance phase. According to this logic, we are tapping into processes of long-term memory and spontaneous retrieval in versions of the paradigm with no processing costs to ongoing behaviour.

Three problems confront this logic. First, results have not been consistent – despite the general finding of a reduced cost when the cue for the PM task is focal compared to a non-focal cue (Anderson, Strube, & McDaniel, 2019; cue focality is high when there is a high overlap between the processing required to complete the ongoing task and the processing to detect a PM target). Some have found reduced processing costs when the instruction describes a cue with high specificity (Ellis & Milne, 1996), but others have found the reverse when specificity is interacting with importance (Marsh, Hicks, & Cook, 2005). Some have found an absence of cost when cues are focal (Scullin et al., 2010), but other studies have found some cost compared to the control condition (Cohen et al., 2012; Cona et al., 2013; McDaniel et al., 2013). The confusing results have motivated researchers to look for indications of spontaneous retrieval in situations where we would have prior reasons to think that participants are not maintaining any PM task in working memory. For instance, situations where participants have been instructed that the PM task is not yet effective (Knight et al., 2011) or no longer effective (Scullin & Bugg, 2013). However, it is not clear how one should interpret these indirect measures (Smith, 2016).

Second, the cost to the ongoing behaviour might not be a consequence of general resource limitations of working memory related processing. The general assumption often made in research on prospective memory is that maintaining an intention active in working memory and actively monitoring for cues exhaust the pool of available resources, so there is less left for the ongoing task. However, if Heathcote and colleagues' "delay theory" is correct (Heathcote, Loft, & Remington, 2015), longer response times in the ongoing task is a result of a strategic delay to gather more evidence. In the PM paradigm, the participant has two competing tasks (the ongoing and the PM). If there is a risk of confusing the two, the participant delays responding and collects more evidence about which response to execute. Participants might proactively adjust evidence thresholds and reactively influence the evidence accumulation rate (Strickland et al., 2018). According to this interpretation, the absence of a cost to the ongoing task is not an indication of spontaneous retrieval from long-term memory but rather an indication that the cue for the PM response is not easily confused with the cue for the response in the ongoing task, i.e., the participant can safely set the thresholds low. Note that like Heathcote and colleagues' delay theory, according to MIS, increased response latencies on an ongoing task do not necessarily indicate a resource demanding form of monitoring. MIS could explain the latency not only by decision thresholds but also by the number of intentions entering the competition for selection.

Third, fMRI studies have not consistently supported the claim that the PM paradigm can study long-term memory for intentions. If the absence of a cost to the performance in the ongoing task in conditions with specific and focal cues was a reliable indication that an agent was not maintaining an intention or task instruction in working memory, then we should not expect activation of working memory related prefrontal areas above baseline activation in tasks with focal cues. According to one prominent model of working memory and cognitive control

(Braver, 2012), when an agent maintains her intention in working memory, she is engaged in "proactive control" and we should expect a maintained above base-line level of activation of prefrontal cortex. By contrast, if an agent forms a future-directed intention to be retrieved on encountering a specific cue, she is engaged in "reactive control" and we should expect transient levels of activation of prefrontal cortex in response to the cue. An important problem for the models of prospective memory closely tied to the PM paradigm is the finding from fMRI experiments that relevant areas of the prefrontal cortex often are activated during the maintenance phase of the PM paradigm (Cona et al., 2016; Gilbert, 2011; Momennejad & Haynes, 2013). In other words, researchers have difficulties finding the type of transient activity one would have expected if the intention were stored in long-term memory and spontaneously retrieved upon encountering the cue.

In sum, this gives us some reason to be sceptical about the use of the PM paradigm to study task-sets in long-term memory. Furthermore, even if we accept that the paradigm successfully encourages a processing strategy according to which the PM task-set is stored in long-term memory, the PM paradigm is still unable to study the selection of a task-set from *a plurality* of task-sets in long-term memory. In the paradigm, participants are required to switch between a frequent ongoing task and an infrequent PM task. The ongoing task is active in working memory. So, if the PM task is represented in long-term memory, the paradigm only requires participant to store one task-set in long-term memory. Therefore, the paradigm cannot be used to study the selection from a plurality of standing task-sets. Given this conclusion, MIS and WM-CM do not make conflicting predictions about behaviour in the PM paradigm. In terms of the model curves (Figure 2), MIS and WM-CM both predict a curve corresponding to a single intention (Figure 2A).

It could be objected that some studies have operated with multiple PM targets (typically, six or eight, Cohen et al., 2012; Humphreys et al., 2020) or have directly manipulated the number of PM targets (1-6 targets, Cohen, Jaudas, & Gollwitzer, 2008). One might argue that it is difficult to see how participants could hold all six or eight PM targets in working memory while working on the ongoing task. However, this does not necessarily require that multiple PM task-sets are held in long-term memory. In this modified version of the PM paradigm, all PM targets require the same response, and only one PM target is presented at a time. A single general intention ("I must remember to press F7 when certain targets are presented") could be held in working memory and the participants could rely on recognition memory to determine what targets should cue them to perform their action (see Strickland et al., 2021).

The conclusion that the PM paradigm is unable to study the selection of a task-set from a plurality in long-term memory has consequences for the relevance of formal models of prospective memory. Even if a model's background assumptions would allow that task-sets are represented in long-term memory, models like Heathcote and colleagues' linear ballistic accumulator model of PM (Heathcote et al., 2015, and extensions in Strickland et al., 2018 and Boag et al., 2019; see also Boywitt & Rummel, 2012 and Horn, Bayen, & Smith, 2011 for alternative drift diffusion models) have so far only been used to model data only from the event-based PM paradigm.

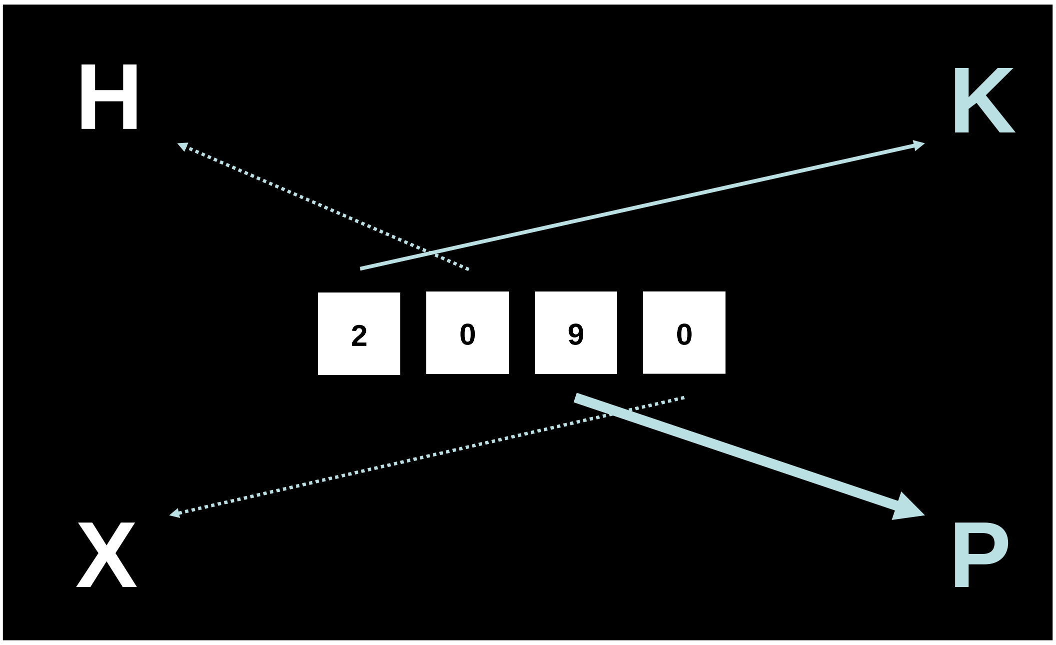

Figure 3. The stimulus and task setup in the new Intention Selection Paradigm. The central four squares represent the four cues with four reward values of 2, 0, 9, and 0 points. The four letters in the corners are the letters of the four possible task-sets/intentions. The arrows in the picture indicate the association between the four cues and the locations of letters to be reported in the four tasks. Reporting the letter P in the lower left corner is optimal and will be rewarded by 9 points and reporting the letter K in the top left corner will be suboptimal but result in a reward of 2 points. Reporting either of the two remaining letters will not be rewarded.

6. A new paradigm of intention selection from long-term memory

How may MIS be tested against WM-CM? In Section 5, we outlined the limitations in already well-established paradigms of task-switching and prospective memory. We argued that the task-switching paradigms are problematic if we want to test predictions about the role of long-term memory because of two limitations. (a) The procedures are not designed to rule out the possibility that the task-sets are maintained in working memory and (b) the procedures cannot be used to study selection from a plurality of task-sets in long-term memory – usually only two task-sets are used in the paradigms, where at least one is always active in working memory. Results from the task-switching paradigms systematically underdetermine the choice between MIS and WM-CM. Given that MIS postulates an additional form of long-term memory, simplicity considerations might move one to adopt WM-CM.[2]

Two crucial differences between MIS and WM-CM are capacity limitation and biased competition. According to MIS, the selection of a task-set from a plurality of potentially relevant task-sets is (1) *capacity limited* in the sense that the higher the number of relevant

[2] For an in-depth description of this type of local but systematic underdetermination by data, see Grünbaum, 2018.

task-sets, the slower the time-course for processing and selecting the optimal task-set (see Equation 2), and (2) *biased competition* between task-sets in the sense that the importance of the optimal task-set relative to the other task-sets determines the probability and time-course of selecting it for performance (see Equation 3). By contrast, WM-CM predicts no capacity limitation and no biased competition because there will always just be one intention or task-set. In a situation where several task-sets are relevant by the same degree of contextual match and varying degrees of importance, MIS therefore predicts performance that could be described by the curves presented in Panels B and C. Given WC-MC only process one intention at any given time, WM-CM predict curves like those in Panel A. WM-CM can only predict the curves like the ones in Panel B and C if the apparent capacity limitations and competition effects can be explained in some other way.

The new paradigm should enable us to investigate the competition between several intentions stored in long-term memory represented by the relative intention weights, $\omega_\chi/\sum\omega$, in MIS (see Equation 1). Imagine that participants overlearn task-cue associations in a training period. Each cue ends up being associated with a separate task-set. We can now have a situation where several cues are presented at the same time. If in each trial we link each cue with a reward value for performing the associated task-set, the selection of the task-set should optimally be governed by the reward values through the intention weights, ω_χ. The practice of each task in the training period and imposed task-rule complexity (the association of each cue with a separate reporting task; see Figure 3) should ensure that task-sets are stored in long-term memory. Woodman, Carlisle, and Reinhart (2013) demonstrated that with task repetitions, the tasks quickly become encoded in long-term memory. Future studies would have to manipulate task-rule complexity in such a way that the complexity of task-rules would exceed working memory capacity.

Here is one suggestion for implementing the new paradigm. Four squares are presented simultaneously as four task cues in the centre of the screen (the four white squares on a line in Figure 3). The cues indicate four possible tasks that the participant can select for execution in a given trial. The task is to report a single letter presented at an associated corner of the screen. That is, each of the four cues are associated with a specific corner in which to report a letter, indicated by arrows in Figure 3. In each trial, the participant has to select a single task (that is, which corner to report a letter from). In each trial, this selection is influenced by linking a cue with a reward value from zero to nine. The letters are presented only very briefly in the four corners to ensure a sufficient level of task difficulty and each of the four letters are drawn without replacement in each trial from all the letters of the alphabet.

The crucial purpose of the experiment will be to manipulate the reward value associated with each cued task and the time between cue presentation and presentation of letters at the four corners of the screen. We can now define optimal performance of a trial as when a participant correctly reports the letter from the corner associated with the highest reward (indicated by the thick arrow in Figure 3) and suboptimal performance as when she correctly reports a letter from a corner associated with a lower reward (i.e., reporting the letter associated with the leftmost cue giving a reward value of two points, indicated by the thin arrow in Figure 3).

Across trials, the number of positively rewarded cues is varied between one to four. The reward values are drawn without replacement, so that all presented cues are associated with a different positive reward. The reward values are expected to motivate the participants to set their intention weights differentially for each of the four task-sets. Thus, manipulation of the reward value corresponds to a manipulation of the relative intention weights, $\omega_\chi/\sum\omega$, in MIS (see Equation 1). This would enable us to test predictions about capacity limitations and competition between task-sets. Finally, the SOA between the presentation of the cues and presentation of the stimuli for the letter report task is systematically manipulated, say, between zero and one second. Manipulation of the SOA would enable us to model the data as a stochastic time dependent process, thus estimating the limited processing capacity of the selection race, C_I, as assumed by MIS (see Equations 2 and 3).

How do MIS and WM-CM conceive of the processing leading to the task performance in this type of paradigm? Both models would assume that processing happens in two successive stages. In the first stage, cues are encoded, and in the second stage, the task-set is selected or constructed. But the two models differ with respect to how they conceive of the two stages.

MIS assumes that the first stage is a parallel encoding of the reward values and location of the cues. According to MIS, all four cues are encoded before the second stage is initiated (exhaustive processing, see Houpt et al., 2014; Townsend & Ashby 1983, Ch. 4). In the second stage, all encoded cues are matched against long-term representations of intention components and thereby individually trigger the selection race between the task-sets for encoding into working memory and subsequent performance. Here the reward-based setting of intention weights is the critical factor effecting task performance. According to MIS, a suboptimal report of a letter is due to a suboptimal task-set winning the selection race. Given MIS, a suboptimal selection of intention could be due to chance factors or the assignment of the "wrong" weight.

In contrast to the MIS interpretation of the task, a participant in the proposed task could be implementing one of at least three different working memory strategies. First, she might maintain the four task-sets in working memory as one super task-set consisting of a long disjunction (if cue *x*, then *a*; or if cue *y*, then *b*; or…). Second, she might maintain four separate task-sets in working memory. Third, she might maintain a general task rule (for instance, "locate the digit with highest value and move attention to associated corner") in working memory and then, given the cue, construct the task-set for each new trial. For the remainder of our discussion, we will assume that the task-rule complexity is sufficiently high to rule out the first and the second working memory strategy.

With this caveat, we can continue. WM-CM also assumes that the first stage is a parallel encoding of the reward values. According to WM-CM, the second stage is initiated already when the first cue is encoded (self-terminating processing, see Houpt et al., 2014; Townsend & Ashby, 1983, Ch. 4). We can think of the first stage as a search for the cue with the highest value. In the second stage, the encoded cue triggers the construction of the corresponding task-set in working memory. In this stage, relative intention weights play no role. Thus, according to WM-CM, suboptimal report of a letter is due to suboptimal visual encoding of the cues, i.e., a cue not having the highest value is incorrectly encoded as having the highest value.

In Figure 2, the predictions of the two models are shown (see Sections 4.2 and 4.3). Panel A indicates the predicted time course of the task-set construction when the highest reward cue is always selected. Panels B and C indicate predicted performance when the encoding of the cues is sometimes suboptimal. Recall that MIS explains suboptimal letter reporting by the selection of the suboptimal task-set, whereas WM-CM explains suboptimal letter reporting by suboptimal encoding of the cues.

Focus now on the first stage with cue encoding. In a situation where participants sometimes perform suboptimally, that is, perform the task not associated with the highest reward value, we would see performance that could be described by panel B and C. In this type of situation, WM-CM will have to explain the performance described by panel B and C by mistakes in the first stage of cue encoding. It could not be in the second stage, since according to WM-CM only one task-set is constructed and maintained at any given time. By contrast, MIS would assume that processing in the first stage is almost error free (no selection of the wrong cue). MIS explains the suboptimal task-performance described in panel B and C by selection of a suboptimal task-set in the second stage of task-set selection. Assuming that the new paradigm described above would provide us with the curves described in Panel B and C, WM-CM would thus predict that the visual selection of the highest digit would be error prone. Therefore, if people are generally accurate at selecting the highest value in an array of digits, it would count against WM-CM.

A number of studies have used paradigms where participants have to search for and report the highest digit in an array of digits (Pashler & Badgio, 1985, Exp. 3; Blanc-Goldhammer & Cohen, 2014, Exp. 2; see also Becker & Pashler, 2002). Pashler and Badgio (1985) investigated the general question of whether visual stimuli are processed in parallel. They reported results from a task where participants named the highest digit in an array of digits measuring reaction times and accuracy. They found evidence for parallel processing of the digits rather than serial processing. Pashler and Badgio suggest a possible parallel model where representations of all possible digits in the stimulus set are activated in parallel and the highest digit representation that is active is reported. Importantly, the results of the experiment supported the assumptions that processing of the digits were both parallel and fast. In their Figure 3 displaying reaction time as a function of display size, the rate of processing of the digits is approximately 30 ms/digit corresponding to a processing rate of 33 digits/s (Townsend & Ashby, 1983, Ch. 4). In several studies, we have found comparable rates of processing of alphanumeric stimuli in whole report and partial report assuming fixed capacity independent parallel processing (Finke et al., 2005; Kyllingsbæk, 2006).

Blanc-Goldhammer and Cohen (2014) investigated the processing architecture and capacity characteristics of quantity comparison of multiple integers (digits). They found strong evidence that integers are encoded, identified, and compared within a capacity unlimited system with a parallel processing architecture. Their Experiment 2 is of particular interest in relation to our comparison of MIS and WM-CM. In the experiment, participants were presented with four digits at the corners of an imaginary square centered at fixation. The digits were presented for 65 ms and patterns masks covering the locations of the digits were presented for 500 ms both before and after the presentation of the digits. In one response condition, participants had to identify the highest digit without time pressure by typing the response on a number pad. The average probability of identifying the highest digit correctly was estimated at a value of 0.80. Given that the digit stimuli were presented with both pre and post masks, the high accuracy and short exposure duration of 65 ms indicate that processing of the highest digit was both fast and accurate.

To conclude, in the situation described by our new paradigm, MIS predicts performance that could be described by the curves presented in Panels B and C, whereas WC-MC can only predict these curves if the apparent capacity limitations and competition effects can be explained in some other way. WM-CM can explain the effects apparent in Panel B and C by claiming that visual cue encoding is error prone. However, data from relevant studies of digit search strongly support the claim that visual processing in stage 1 is accurate. Given the experimental data on digit search, if the curves from our new paradigms look like Panel B and C, MIS is more likely to be a correct model than WC-CM. This would provide us with a reason to think the LTM conception is more likely to be correct than the WM conception of the role of memory in temporally extended agency.

Our aim has been to present our new model of intention selection (MIS), derive a number of testable quantitative predictions, compare them to predictions derived from a competing formal model (WM-CM), and finally suggest a possible paradigm to test the predictions. Existing task-switching paradigms are unable to test the models because they operate with experimental situations where participants are at most processing only a single intention from long-term memory. To test the models against each other, we need a paradigm that is able to manipulate the number of relevant task-sets in long-term memory. In this section, we have exemplified just one possible way in which one might test the model predictions.

7. Conclusion

We have outlined two opposing conceptions of the role of memory in temporally extended agency – as when a person has to remember to make a phone call in the afternoon because, in the morning, she promised she would do so. According to the WM conception, remembering to make the phone call is explained in terms of the construction and maintenance of intentions in working-memory. According to the LTM conception, we should explain the episode in terms of an ability to store intentions in long-term memory. We argued that this type of long-term memory for intentions would be characterised by two features: (1) intentions are stored at a low cognitive cost and (2) a plurality of intentions is stored, so retrieval is a selection process.

On this background, we introduced a new formal model (MIS) of the selection of intentions from Intention Long-Term Memory (ILTM). MIS is a formal version of the LTM conception. We used the formal framework of MIS to define a formal model WM-CM corresponding to the WM conception. MIS and WM-CM make conflicting predictions about how participants will behave in situations with several cues and several relevant task-sets. We can use these conflicting predictions to tell us whether the WM conception or the LTM conception of the role of memory is most likely to be correct.

By selectively reviewing studies of task-switching and prospective memory, we showed that standard experimental procedures are unable to investigate directly these features of long-term memory. Results from task-switching paradigms (standard task-switching procedures and event-based PM procedures) systematically underdetermine the choice between MIS and WM-CM. We therefore proposed a new intention selection paradigm that might allow us to manipulate the number of relevant standing intentions in long-term memory. The proposed intention selection paradigm outlines only one way to test the predictions of MIS and WM-CM. Future studies will have to investigate whether actual data from selection paradigms support one or the other model.

Supplementary Material: Appendix 1

Mathematical derivations of the time course of selection between several intentions.
URL: [https://ars.els-cdn.com/content/image/1-s2.0-S0010027721002365-mmc1.docx]

Acknowledgements

This project owes a special debt to the support of Gordon Logan. We are indebted to Frederick Verbruggen, Ollie Hulme, and Mark Schram Christensen for pushing us to articulate and formalize the Working Memory conception. We are grateful to Steven Blurton for help with the mathematical derivations. We thank members of the CoInAct group for their useful comments on the manuscript.

Funding

This work was supported by the Independent Research Fund Denmark [grant number: DFF – 6107-00058].

References

Allport, A., Styles, E. A., & Hsieh, S. (1994). Shifting attentional set: Exploring the dynamic control of tasks. In C. Umiltá & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious information processing* (pp. 421–452). Cambridge, MA: MIT Press.

Anderson, F., Strube, M., & McDaniel, M. (2019). Toward a Better Understanding of Costs in Prospective Memory: A Meta-Analytic Review. *Psychological Bulletin*, *145*(11), 1053-1081.

Austin, J., & Vancouver, J. (1996). Goal Constructs in Psychology: Structure, Process, and Content. *Psychological Bulletin*, *120*(3), 338-375.

Baddeley, A. (2000). The episodic buffer: a new component of working memory?. Trends in cognitive sciences, 4(11), 417-423.

Becker, M. W., & Pashler, H. (2002). Volatile visual representations: Failing to detect changes in recently processed information. *Psychonomic Bulletin & Review*, *9*(4), 744-750.

Blanc-Goldhammer, D. R., & Cohen, D. J. (2014). Unlimited capacity parallel quantity comparison of multiple integers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1389.

Blurton, S. P., Kyllingsbæk, S., Nielsen, C. S., & Bundesen, C. (2020). A Poisson random walk model of response times. *Psychological Review*, *127*(3), 362-411.

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.

Boywitt, C. D., & Rummel, J. (2012). A diffusion model analysis of task interference effects in prospective memory. *Memory & Cognition*, *40*(1), 70-82.

Brass, M., Liefooghe, B., Braem, S., & De Houwer, J. (2017). Following new task instructions: Evidence for a dissociation between knowing and doing. *Neuroscience & Biobehavioral Reviews*, *81*, 16-28.

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

Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in cognitive sciences*, *16*(2), 106-113.

Broome, J. (2013). *Rationality through reasoning*. John Wiley & Sons.

Bugg, J.M. & Streeper, E. (2019). Fate of suspended and completed prospective memory intentions. In Rummel, J., & McDaniel, M. A. (Eds.). (2019). *Prospective memory*, pp. 44-59. Abingdon: Routledge.

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

Butterfill, S. A., & Sinigaglia, C. (2014). Intention and motor representation in purposive action. *Philosophy and Phenomenological Research*, *88*(1), 119-145.

Christensen, J. H., Markussen, B., Bundesen, C., & Kyllingsbæk, S. (2018). A physiologically based nonhomogeneous Poisson counter model of visual identification. *Journal of experimental psychology. Human perception and performance*. doi: 10.1037/xhp0000539

Cohen, A. L., Jaudas, A., & Gollwitzer, P. M. (2008). Number of cues influences the cost of remembering to remember. *Memory & cognition*, *36*(1), 149-156.

Cohen, A. L., Jaudas, A., Hirschhorn, E., Sobin, Y., & Gollwitzer, P. M. (2012). The specificity of prospective memory costs. *Memory*, *20*(8), 848-864.

Cona, G., Bisiacchi, P. S., & Moscovitch, M. (2013). The effects of focal and nonfocal cues on the neural correlates of prospective memory: insights from ERPs. *Cerebral cortex*, *24*(10), 2630-2646.

Cona, G., Bisiacchi, P. S., Sartori, G., & Scarpazza, C. (2016). Effects of cue focality on the neural mechanisms of prospective memory: A meta-analysis of neuroimaging studies. *Scientific reports*, *6*, 25983.

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.

Einstein, G. O., & McDaniel, M. A. (1990). Normal aging and prospective memory. *Journal of experimental psychology. Learning, memory, and cognition*, *16*(4), 717-726.

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.

Finke, K., Bublak, P., Krummenacher, J., Kyllingsbæk, S., Müller, H. J., & Schneider, W. X. (2005). Usability of a theory of visual attention (TVA) for parameter-based measurement of attention I: Evidence from normal subjects. *Journal of the International Neuropsychological Society: JINS*, *11*(7), 832.

Gibson, J. J. (1941). A critical review of the concept of set in contemporary experimental psychology. *Psychological Bulletin*, *38*, 781–817.

Gigerenzer, G., & Selten, R. (Eds.). (2002). *Bounded rationality: The adaptive toolbox*. MIT press.

Gilbert, S.J. (2011). Decoding the content of delayed intentions. *Journal of Neuroscience*, 31, 2888-2894.

Gollwitzer, P. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist*, *54*(7), 493-503.

Grange, J., & Houghton, G. (Eds.). (2014a). *Task Switching and Cognitive Control*. Oxford: Oxford University Press.

Grange, J.A., & Houghton, G (2014b). Task switching and cognitive control: An introduction. In J.A. Grange & G. Houghton (Eds.), *Task switching and cognitive control*. New York, NY: Oxford University Press.

Grünbaum, T. (2018). The Two Visual Systems Hypothesis and Contrastive Underdetermination. *Synthese*. https://doi.org/10.1007/s11229-018-01984-y

Grünbaum, T., & Kyllingsbæk, S. (2020). Is Remembering to do a Special Kind of Memory? *Review of Philosophy and Psychology*, *Volume 11*, *Issue 2*, 385-404.

Heathcote, A., Loft, S., & Remington, R. W. (2015). Slow down and remember to remember! A delay theory of prospective memory costs. *Psychological Review*, *122*(2), 376.

Hommel B. (2006), How we do what we want: A Neuro-cognitive perspective on human action planning. In: Jorna R.J., Wezel, van W., Meystel A. (Eds.) *Planning in intelligent systems: Aspects, motivations and methods*. New York: John Wiley & Sons. 27-56.

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.

Houpt, J. W., Blaha, L. M., McIntire, J. P., Havig, P. R., & Townsend, J. T. (2014). Systems factorial technology with R. *Behavior research methods*, *46*(2), 307-330.

Humphreys, M. S., Li, Y. R., Burt, J. S., & Loft, S. (2020). How semantic processing affects recognition memory. *Journal of Memory and Language*, *113*, 104109.

Kadohisa, M., Watanabe, K., Kusunoki, M., Buckley, M. J., & Duncan, J. (2020). Focused Representation of Successive Task Episodes in Frontal and Parietal Cortex. *Cerebral Cortex*, 30(3), 1779–1796.

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching—A review. *Psychological Bulletin*, *136*, 849–874.

Knight, J., Meeks, J., Marsh, R., Cook, G., Brewer, G., & Hicks, J. (2011). An Observation on the Spontaneous Noticing of Prospective Memory Event-Based Cues. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(2), 298-307.

Kyllingsbæk, S. (2006). Modeling visual attention. *Behavior Research Methods*, *38*(1), 123-133.

Kyllingsbæk, S. (2016). Visual Processing Capacity. In *Mathematical Models of Perception and Cognition Volume II* (pp. 50-64). Routledge.

Kyllingsbæk, S., Markussen, B., & Bundesen, C. (2012). Testing a Poisson counter model for visual identification of briefly presented, mutually confusable single stimuli in pure accuracy tasks. *Journal of Experimental Psychology: Human Perception and Performance, 38*, 628-642. doi: 10.1037/a0024751

Kvavilashvili, L., & Ellis, J. (1996). Varieties of intention: Some distinctions and classifications. In M Brandimonte, G. O. Einstein, & M. A. McDaniel (eds.), *Prospective memory: Theory and applications*, pp. 183-207. Mahwah, NJ: Lawrence Erlbaum.

Logan, G. D., & Bundesen, C. (2003). Clever homunculus: Is there an endogenous act of control in the explicit task-cuing procedure? *Journal of Experimental Psychology: Human Perception and Performance, 29,* 575–599.

Logan, G. D., & Bundesen, C. (2004). Very clever homunculus: Compound stimulus strategies for the explicit task-cuing procedure. *Psychonomic Bulletin & Review, 11,* 832–840.

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

Marsh, R., Hicks, J., & Cook, G. (2005). On the Relationship Between Effort Toward an Ongoing Task and Cue Detection in Event-Based Prospective Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(1), 68-75.

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

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., LaMontagne, P., Beck, S. M., Scullin, M. K., & Braver, T. S. (2013). Dissociable Neural Routes to Successful Prospective Memory. *Psychological Science,* 1791-1800.

Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1423–1442.

Meiran, N. (2014). *The task-cuing paradigm: A user's guide.* In J. A. Grange & G. Houghton (Eds.), *Task switching and cognitive control* (p. 45–73). Oxford University Press.

Mele, A. R. (1992). *Springs of action: Understanding intentional behavior*. Oxford: Oxford University Press.

Mele, A. R. (2009a). Intention and Intentional Action. In B. McLaughlin and A. Beckermann, eds. *Oxford Handbook of Philosophy of Mind*. Oxford: Oxford University Press, pp. 691-710.

Mele, A. R. (2009b). *Effective intentions: The power of conscious will*. Oxford: Oxford University Press.

Momennejad, I., & Haynes, J. D. (2013). Encoding of prospective tasks in the human prefrontal cortex under varying task loads. *Journal of Neuroscience*, *33*(44), 17342-17349.

Monsell, S. (2003). Task switching. *Trends in cognitive sciences*, *7*(3), 134-140.

Monsell, S., Sumner, P., & Waters, H. (2003). Task-set reconfiguration with predictable and unpredictable task switches. *Memory & Cognition, 31,* 327–342.

Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press

Oliva, A. (2005). Gist of the scene. In *Neurobiology of attention* (pp. 251-256). Academic press.

Pacherie, E. (2008). The phenomenology of action: A conceptual framework. *Cognition, 107*(1), 179-217.

Pashler, H. (1989). Dissociations and dependencies between speed and accuracy: Evidence for a two-component theory of divided attention in simple tasks. *Cognitive Psychology*, *21*(4), 469-514.

Pashler, H., & Badgio, P. C. (1985). Visual attention and stimulus identification. *Journal of Experimental Psychology: Human Perception and Performance*, *11*(2), 105.

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.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124,* 207–231.

Rummel, J., & McDaniel, M. A. (Eds.). (2019). *Prospective memory*. Routledge.

Sacerdoti, E. D. (1974). Planning in a hierarchy of abstraction spaces. *Artificial intelligence*, *5*(2), 115-135.

Searle, J. R. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge: Cambridge University Press.

Schneider, D. W., & Logan, G. D. (2014). Tasks, task sets, and the mapping between them. In J. A. Grange & G. Houghton (Eds.), *Task switching and cognitive control* (pp. 27-44). New York: Oxford University Press.

Scullin, M. K., & Bugg, J. M. (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., McDaniel, M. A., & Einstein, G. O. (2010). Control of cost in prospective memory: Evidence for spontaneous retrieval processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 190-203.

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), 736-749.

Simon, H. A. (1957). *Models of Man*. New York: John Wiley & Sons.

Smith, R. E. (2003). The cost of remembering to remember in event-based prospective memory: investigating the capacity demands of delayed intention performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(3), 347.

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.

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., Hunt, R., McVay, J., & McConnell, M. (2007). The Cost of Event-Based Prospective Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 734-746.

Spector, A., & Biederman, I. (1976). Mental set and mental shift revisited. *Journal of Psychology, 89,* 669–679.

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.

Strickland, L, Heathcote, A, Humphreys, M. S., & Loft, S. (in press). Target learning in event-based prospective memory.

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.

Sudevan, P., & Taylor, D. A. (1987). The cuing and priming of cognitive operations. *Journal of Experimental Psychology: Human Perception and Performance*, *13,* 89–103.

Sumner, P., & Ahmed, L. (2006). Task switching: The effect of task recency with dual- and single-affordance stimuli. *Quarterly Journal of Experimental Psychology, 59*, 1255–1276.

Townsend, James T., and F. Gregory Ashby. (1983). *Stochastic modeling of elementary psychological processes*. Cambridge, MA: Cambridge University Press.

Vandierendonck, A., Liefooghe, B., & Verbruggen, F. (2010). Task Switching: Interplay of Reconfiguration and Interference Control. *Psychological Bulletin*, *136*(4), 601-626.

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.

Vidal, J. R., Gauchou, H. L., Tallon-Baudry, C., & O'Regan, J. K. (2005). Relational information in visual short-term memory: the structural gist. Journal of vision, 5(3), 244-256.

Woodman, G. F., Carlisle, N. B., & Reinhart, R. M. (2013). Where do we store the memory representations that guide attention?. *Journal of Vision*, *13*(3), 1-17.