Remembering as a mental action[*]

Forthcoming in *New Directions in the Philosophy of Memory*
(ed. K. Michaelian, D. Debus & D. Perrin). Routledge. Please cite print version.

Santiago Arango-Muñoz
G.I. Conocimiento, Filosofía, Ciencia, Historia y Sociedad
Instituto de Filosofía
Universidad de Antioquia

Juan Pablo Bermúdez
Facultad de Ciencias Sociales y Humanas
Universidad Externado de Colombia

Abstract: Many philosophers consider that memory is just a passive information retention and retrieval capacity. Some information and experiences are encoded, stored, and subsequently retrieved in a passive way, without any control or intervention on the subject's part. In this paper, we will defend an active account of memory according to which remembering is a mental action and not merely a passive mental event. According to the reconstructive account, memory is an imaginative reconstruction of past experience. A key feature of the reconstructive account is that given the imperfect character of memory outputs, some kind of control is needed. Metacognition is the control of mental processes and dispositions. Drawing from recent work on the normativity of automaticity and automatic control, we distinguish two kinds of metacognitive control: top-down, reflective control, on the one hand, and automatic, intuitive, feeling-based control on the other. Thus, we propose that whenever the mental process of remembering is controlled by means of intuitive or feeling-based metacognitive processes, it is an action.

---

[*] This research was supported by the "Metacognitive Feelings and the Cognitive Phenomenology Thesis" research project funded by the CODI, Acta 01 de 2016, Universidad de Antioquia, and by the John Templeton Foundation's "Getting Better at Simple Things: Understanding and improving vigilant control" project.

1. Introduction

There has been a recent shift in our understanding of episodic memory. Traditionally conceived as the capacity to store and retrieve information from our personal past, episodic memory is now thought of as a particular form of a broader cognitive function: that of imagining, or mentally simulating, ego-centric events, whether they belong to the past or the future, whether actual, hypothetical, or counter-factual. Understanding episodic memory as part of this more general capacity for 'mental time-travel' allows us to account for why it often produces not a reliable reconstruction of the specific details of the past, but rather a bare-bones reconstruction of past situations that diverge from the remembered events.

In this paper we seek neither to defend this reconstructive conception of episodic memory, nor to argue that it is empirically better grounded than the traditional, preservative account (These tasks have been performed by De Brigard (2014a), Michalian (2016), and Schacter & Addis (2007), and we refer the interested reader to these works). There are at least two ways to understand the constructive theory of memory: as episodic hypothetical thought (De Brigard, 2014a), and as episodic imagination (Michaelian, 2016), but we need not decide among them here. Rather, we assume the imaginative reconstructive conception (but analogous considerations should apply to the other version), and ask a further philosophical question: does remembering constitute a mental action? Is the imaginative reconstruction, or mental simulation, involved in episodic memory, an action? If so, what kind of agentive processes and mechanisms are at stake in remembering?

We argue that in some cases remembering is indeed a mental action. But in order to argue for this we have to face objections from thinkers who worry that the "ballistic" nature of mental content production leaves no room for agency and control in imaginative processes. In particular, Strawson (2003) has argued that although there are some cases of mental actions, they take place prior to, and merely 'set the stage' for, the emergence of mental content. I may agentively try to remember the first time I read *The Myth of Sisyphus*, and these efforts to bring a certain mnemonic content to mind may count as mental actions, but the actual imaginative reconstruction of the situation in which I first encountered that text (images of the place, the time, the people around me, etc.) is not under my agentive control. It occurs

spontaneously, in the sense of 'involuntarily'; its production mechanisms are outside of my awareness; I could not control them, as demonstrated by the fact that if I intend to *not* remember the first time I read Camus' text, I would still end up bringing (at least some of) these mental contents to mind.

This is the objection from mental ballistics: imagination is a ballistic mental process, and given that episodic memory is a kind of imagination, episodic remembering is also a ballistic process that cannot be a mental action.

This essay argues against the mental-ballistics interpretation of reconstructive episodic memory. We claim that the mnemonic processes that constitute episodic reconstruction are not necessarily ballistic, but can count as instances of mental action. To show this, (2) we distinguish between two kinds of agentive control: reflective and automatic. The latter is often taken to be impossible, because automaticity and control are often considered opposites; but we show that there are cases of automatic *and* controlled mental processes. Then we argue that (3) there are two levels of metacognition that correspond to the reflective and automatic control processes distinguished above, and that automatic, feeling-based metacognition is a control process that can be applied to mental processes like remembering. Lastly, (4) by looking at the evidence for remembering as an imaginative reconstruction, we argue that this process is often characterized by particular forms of feeling-based metacognitive control, and that the presence of this control process reveals the agentive nature of episodic memory. Remembering is often a mental action, because the imaginative reconstruction processes that constitute it are imperfect, and therefore need some kind of control. To fill this need, our agentive involvement is frequently needed to manage and structure the reconstructive processes, and thus guarantee that we produce good memories as an outcome of the episodic reconstruction. When this happens, episodic reconstruction is constrained and guided by a process of feeling-based metacognitive control.

2. Mental ballistics, automaticity, and control

The idea of memory as reconstruction seems to imply some kind of activity: remembering requires bringing the multiple traces of the remembered experience

together in a mental simulation of the event (De Brigard, 2014b; Michaelian, 2016). Although this looks *prima facie* like an agential process, there is a difficulty: a number of lines of cognitive science research have shown that most of the mental processes that shape our daily lives take place automatically: they are fast, associative, working-memory-independent, and therefore apparently produced outside of effortful cognitive control (for general overviews see Evans (2010b) and Kahneman (2011)).

Thus, although the reconstructive account suggests that memory is an agentive process, it still has to confront the "threat of automaticity" (Wu, 2013): although remembering is a reconstruction, this reconstruction seems to take place automatically without the subject's agential intervention. Wayne Wu remarks that "automaticity is what makes decisions about mental agency controversial" (Wu, 2013, p. 244). Most philosophers consider memory to be a more or less automatic capacity. In a recent paper, for example, Andy Clark claims that "ordinary biological memory, for the most part, functions in a kind of automatic, subterranean way" (Clark, 2015). If this is so, then calling 'remembering' an action is highly controversial.

The core philosophical issue here is that automatic processes seem to be by their very nature non-agential. After all, we tend to call 'automatic' those processes that are invariably triggered by the same kind of stimuli and respond to them in systematically the same way. They are rigid, difficult or impossible to correct, and insensitive to novel evidence. In a word, automatic mental processes seem to be *ballistic*: they "run to completion once triggered and cannot be aborted in midcourse" (Stanovich, 2004, p. 39). Add to this what Wu calls the "simple connection" between automaticity and control: "automaticity implies *the absence of control* […] *by the subject*" (2013, p. 246). This traditional view has been questioned empirically, but it retains a powerful hold over our theoretical intuitions. The simple connection entails that agentive control consists in the deployment of top-down attention and cognitive effort toward the attainment of a goal, and that automatic processes, being fast and effortless, are devoid of agentive control and can participate in it only to the extent that they are subsumed under higher-order, cognitively effortful mental processes.

According to this view, if remembering was to be considered an action, the subject would have to concentrate, pay attention, avoid distracting elements, try mnemonic strategies intentionally, gather cues…, and all this would imply a certain

amount of non-automatic attention, consciousness, and mental effort, even if not particularly "strenuous" (Mele, 1997). But even if this happens, even if the preparatory ground-setting steps are cognitively effortful and controlled, the very process of imaginative reconstruction seems to work spontaneously and sub-personally: when I remember what I had for dinner last night, I cannot really *try to remember:* the memory traces are assembled and presented as if by magic; all I can do at the personal level is *try to bring it about that I remember*, conjuring the content, and wait for it to arrive (Mele, 2009).[1] Thus, if automatic processes are ballistic, there is nothing particularly agential about them.

2.1. The mental ballistics argument: from imagination to memory

Given this view of automatic processes as ballistic, Strawson concludes that the space for action in the mental is rather small: "the role of genuine action in thought is at best indirect. It is entirely *prefatory*, it is essentially—merely—*catalytic*" (2003, p. 236). The mental phenomena that can be properly called actions are those that 'set the stage' for the emergence of mental content, but not the content's emergence itself. Setting myself up for tackling a problem may be a mental action; the actual solution's appearance in my mind is not. Bringing my attention back to a task after I have been distracted may be a mental action; the subsequent steps involved in completing the task probably are not.

This also applies in the case of imagination. If I ask you to imagine a pink elephant gracefully walking on a rainbow, you can do this immediately and intuitively, in a way that suggests agentive control: you can picture it big or small, you can make it dance, smile or wave its hat. This sense of agentive control over imagination is, however, set in a different light if I ask you *not* to imagine a pink elephant at all. Now it is harder to make imagination comply: you may *try to bring it about that* you do not imagine the elephant (by occupying your mind with something

[1] This distinction between "trying to *X*" and "trying to bring it about that *X*", plus the claim that one can do only the latter with respect to memory, led Mele to revise his earlier view (Mele, 1997) and conclude that remembering is never a mental action (Mele, 2009).

else), but you cannot directly *try to* not imagine the elephant. This reveals that the actual imaginary content production is a rather ballistic and rigid process that takes one input and automatically produces content in a way that is not responsive to the agent's intention, its structure and content (perhaps the experience of control in the first case was due to the ease with which the content was produced, rather than with proper intentional control). Thus, the imaginative production of mental content seems not to be a case of mental action. If so, the only agentive part of the process would be the initial mental stage-setting.

Applied to imagistic production, the argument from mental ballistics goes like this:

i) If M is a mental action, then I can control the occurrence of M by *both* intentionally trying to perform M *and* intentionally trying not to perform M.

ii) I cannot control the occurrence of P (the imaginative production of a mental content) by intentionally trying not to perform P.

iii) Therefore, P is not a mental action.

Now, if episodic memory is a kind of imagination, would the same argument apply to it? If we ask you to remember what your childhood bedroom looked like, it seems like you can exert swift, even effortless control over this process, focusing on different aspects of the room (the floor, the positions of the bed and other furniture, the lighting…). But if we ask you *not* to remember your childhood bedroom, the recollection process becomes harder to control. Mnemonic contents, like imaginary ones, seem to be produced automatically after the reception of a relevant stimulus, and therefore, ballistically and outside of the subject's agentive control.

2.2. The simple-connection reply to the mental ballistics argument

Here is a way to reply to these arguments: content generation may be automatic and ballistic, but it is nevertheless a constitutive part of mental action whenever the automatically-generated content is triggered by, and responsive to, the agent's occurrent intention. Call this the 'simple-connection reply', since it argues for imagination, deliberation, and the like being mental actions, but does so without

putting the simple connection (i.e. automatic processes are not controlled) into question.

Briefly, the strategy is as follows (Wu, 2013): mental action requires the production of a specific mental content out of many possible ones (in trying to remember my childhood bedroom, I could end up generating images of my current bedroom, other rooms of my childhood home, someone else's childhood bedroom, and so on); and so mental action requires selecting the right content to be generated. Thus, a certain mental process counts as an action if the produced mental content corresponds to the intention's representational content, since this implies that the agent has selected the right content, i.e. has directed her attention to the proper content. This can occur only if the agent's occurrent intention plays a top-down causal role in directing attention toward the right content. And to the extent that her intention plays this causal role, this process is an instance of agentive control, regardless of whether some of its sub-processes are automatic and ballistic.

We think, however, that this simple-connection strategy falls short of fully answering the problem of mental ballistics, for two reasons. First, although we agree that intentions often play this top-down role, the account so far does not explain how intentions do precisely that. How does intention's capacity to coordinate and structure automatic processes work? In other words, what makes automatic processes—otherwise blind, ballistic and unintelligent—responsive to the intention's content, and able to yield adequate results? Although the top-down causal role of intentions should certainly be a part of the story of mental agency, there is another part of it that is missing, i.e. the part that makes automatic mental content-generation processes susceptible to being recruited by higher-order processes.

And second, if we have no story to tell about the agent's role in guiding her automatic processes, then we must still accept Strawson's view that the agent generates an intention, holds it in mind, and then simply waits for the automatic processes—over which she has no control—to generate the relevant content. This fails to agree with our phenomenology of mental action: not only do we experience that we can bring the relevant mnemonic content to mind; we also feel like remembering is sometimes easier, sometimes harder; we feel that a given memory is more or less accurate; sometimes we feel that we can remember something if we try harder,

whereas other times we simply know that we will not be able to remember, no matter how hard we try. All of these phenomena suggest that there is more to agentive mental control than simply holding an intention in mind and waiting for the content. They suggest that we control not only the top-down initiation of remembering, but also (aspects of) the bottom-up production of mental content.

If this is right, and if content production is an automatic process, then there may be a more thorough reply to the mental-ballistics argument than the simple-connection account. And we think that we should let go of the simple connection because there is empirical evidence that some automatic mental processes are *also* controlled.

2.3. Can automatic processes be controlled?

Psychological research has traditionally considered the concepts of automaticity and control as opposites.[2] One of the strongest traditional pieces of evidence for the uncontrollability of automatic processes comes from Stroop-type effects, where unattended dimensions of a stimulus interfere with the attention-demanding task that subjects attempt to perform (MacLeod, 1991; Stroop, 1935). Findings like these suggest that automatic processes are ballistic: the relevant input triggers them almost invariably, and, once activated, they run rigidly to completion. If automatic processes are ballistic, they are not in themselves controlled or controllable.

We present now some evidence that challenges the ballistic interpretation, showing that some automatic processes do not respond invariably to the same stimulus, or run rigidly to completion once triggered, but rather display contextual and normative sensitivity.

---

[2] The picture has become much more complicated than a duality of opposites. However, the original description of the concepts as opposites (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977) still bears influence on contemporary research, and is therefore worth taking as a starting point for conceptual analysis.

Automatic processes and context sensitivity

The traditional view that automatic processes are reflex-like, activated by the mere presence of a given input, carries its influence to this day (Bargh, Chen, & Burrows, 1996; Gendler, 2008a).[3] And yet, no sensible account, no matter how mechanistic, could deny the power of context. Stroop effects themselves have been found to be modulated by contextual features, like how many dissonantly coloured characters each word contains, the participant's direction of attention, and the goals of the task. This suggests that the context that triggers the activation of a given automatic process can include not only external features of the environment, but also the agent's own cognitive dispositional and occurrent states. In fact, there is ample evidence that, contrary to the ballistic interpretation, automatic process activation is affected by "where the current focus of conscious attention is, what the individual was recently thinking, or what the individual's current intentions or goals are" (Bargh, 1997, p. 3).[4]

The conditionality of automaticity is so broad in scope that conceiving automatic processes as ballistic is a misdescription. If reflexes are indeed triggered by the mere presence of the relevant stimulus, then this is a way in which (at least some) automatic processes are *unlike* reflexes, since in many cases there is nothing like a clearly identifiable stimulus whose mere presence almost invariably activates the same automatic process. The latter, on the contrary, are sensitive not only to a specific triggering condition, but also to the agent's motivational states and occurrent goals.

---

[3] Schneider and Shiffrin defined an automatic process as an activation of a series of neural nodes in which "[t]he sequence of nodes (nearly) always becomes active in response to a particular input configuration, where the inputs may be externally or internally generated and include the general situational context" (Schneider & Shiffrin, 1977, p. 2). The view has remained influential: "The essence of TASS [=intuitive] subprocesses is that they trigger whenever their appropriate stimuli are detected, that they cannot be selectively 'turned off'" (Stanovich, 2004, p. 52). In a review of literature on automaticity in social behaviour, Bargh et al. (1996, p. 252) find that "[r]ecent research has shown that attitudes and other affective reactions can be triggered automatically by the mere presence of relevant objects and events, so that evaluation and emotion join perception in the realm of direct, unmediated psychological effects of the environment."—a view that Gendler (2008a, p. 644) quotes approvingly.

[4] For the context-dependence of Stroop effects, see Francolini & Egeth (1980), Kahneman & Henik (1981), and Besner & Stolz (1999). These studies contributed to the rejection of the traditional view that automaticity is attention-independent (LaBerge & Samuels, 1974; Posner & Snyder, 1975; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977).

Automatic error detection and correction

Several researchers have proposed that automaticity's core trait is its being *unstoppable once initiated*. This seems phenomenologically correct: when we see a familiar word we cannot help but read it, and when we see a person's face we cannot but recognize it as such.[5]

It thus seems that automatic processes are *recalcitrant to available evidence*, so that their behavioural output cannot be cancelled or modified even when we have clear evidence for its inappropriateness. Gendler (2008a, 2008b) mentions examples like the fear that even wise persons feel when hanging from a precipice despite being certain that the cage they are suspended in is completely safe, and the unavoidable disgust caused by the prospect of eating delicious, yet feces-shaped, chocolate fudge. Such epistemic insensitivity is yet another reason to consider automatic processes uncontrollable, since they cannot be corrected mid-performance on the basis of updated information.

Yet phenomenology *also* suggests that automatic processes are often self-correcting. When you over-squeeze a plastic cup, you immediately readjust the strength applied by your hand. As you walk, run, skate, cycle, etc., you automatically adjust your posture on the basis of visual and vestibular cues, in ways so complex that explicit calculation would be unable to specify, relying instead on automatic processes acquired through practice. In fact, automatic error-correction is so fast and efficient that oftentimes we are completely unaware of its occurrence. Koch & Crick (2001, p. 893) discuss a study in which participants must move their eyes and fingers rapidly toward an appearing light at their visual field's periphery. They reliably do this even

[5] For Shiffrin & Schneider, "[s]ome automatic processes may be initiated under subject control, but once initiated all automatic processes run to completion automatically" (Shiffrin & Schneider, 1977, p. 160). Norman and Shallice (1986) consider that, since automatic processes run to completion once triggered, any kind of error-correction requires deliberate attentional control. Bargh refers to this as automaticity's "autonomy" (1992, p. 186), and suggests that it is its one core feature. For Stanovich, "TASS [i.e. intuitive] processes cannot be turned off or interfered with by central systems. Their operation is obligatory when triggered by the relevant stimuli […]. TASS processes tend to be ballistic—they run to completion once triggered and cannot be aborted in midcourse" (Stanovich, 2004, p. 39). (See also Logan & Cowan, 1984; Moors & De Houwer, 2006, pp. 301–302).

if the light moves a bit to the left or right as their eyes move towards it, and interestingly they do not report the light having moved. Thus, automatic processes can display an ability to adapt to enviromnental changes, to detect and resolve tensions present within the dynamic stream of associative activity (Brownstein & Madva, 2012; Rietveld, 2008). You probably just performed an automatic error correction in reading this paragraph, by reading 'environmental' instead of 'enviromnental'.

This automatic normativity is not proper to bodily action: it is also present in mental action. In Walsh and Anderson's (2009) paradigm, subjects were confronted with multiplication problems which they could solve using one of two strategies: mentally calculating or using a calculator machine. In selecting the strategy subjects had also to consider a delay of the calculator machine (4 seconds). There were three types of problems: easy, intermediate, and difficult; and two conditions: with calculator delay and without delay. The rate time, the accuracy and the cursor movement from the starting place to the calculator or to the answering box were recorded. Participants quickly initiated a movement corresponding to an initially favored strategy, and then decided whether to complete the problem using that strategy (or shift to another) by conducting a more thorough evaluation while moving the cursor from one place of the screen to another. They often redirected the initial movement of the cursor from one strategy to the other showing "imperfect sensitivity to the current problem"; but then the "commitment to a specific strategy, which occurred later, reflected nearly perfect sensitivity to the profitability of mental and calculator solutions" (Walsh & Anderson, 2009, p. 345). Participants showed an adaptive behavior by performing mental calculations less frequently as the difficulty increased, and more frequently in difficult cases when the calculator was delayed. Strategy selections were extremely fast and depended on the interaction between problem difficulty and calculator responsiveness.

Moreover, there's also fast automatic error detection in mental action. In a recent study, Fernández et al. (2016) designed an experiment to test subjects' awareness of their errors in fast mental calculation. Participants were presented with a triplet and they had to estimate whether the number in the middle was the arithmetic mean of the two other numbers (e.g., 2 4 8) by a Yes/No answer. Participants were instructed to press the Yes/No buttons as fast as possible and then report whether they

had a feeling of error as fast as possible using again the Yes/No buttons. Time pressure was exerted to prevent reflective control and analytical thought. Interestingly, the feeling of error reports were strongly correlated with actual arithmetic errors; in other words, subjects reported having a feeling of error mainly when they had actually committed an error. Additionally, the experimenters tested participants' confidence in their answers when they did not report feelings of error. Surprisingly, in these cases participants reported less confidence for wrong answers than for right ones, suggesting that in cases where subjects did not have a reportable feeling they still had a lower-level awareness of their errors (the feeling of error was in 'fringe consciousness' (E. Norman, Price, & Duff, 2010)).

So we have automatic error-detection systems, which sometimes produce an explicit feeling of error. Add to this that we are able to not only detect, but also correct errors automatically. In a study that compared people's abilities to correct, report, and recall errors (Rabbitt, 2002), participants were instructed to look at a screen split into four squares, and press the corresponding button when a dot appeared in each square. Group 1 was instructed to immediately correct the mistakes they made during the task; group 2 was instructed to press a fifth button every time they had a feeling of error; group 3 was randomly interrupted and asked whether they remembered having made a mistake in their last three responses; and group 4 was told to simply ignore all errors. The stimulus' duration (the Response Signal Interval, or RSI) varied randomly throughout the task between 150ms and 1s.

The first relevant finding was that for all participants the response that followed a mistake was slower than the previous ones. This slowing occurred even when they were unable to report or recall the errors. Further, participants in group 1 were remarkably fast and accurate in correcting their errors: throughout all RSI's, errors were on average faster than correct responses, and error corrections were on average *even faster* than errors. Additionally, participants displayed much more accuracy in error correction than in error report or error recall. Across all conditions, it took participants much longer to report an error than to correct it, and failed to report errors much more often than they failed to correct them. Given that error correction can occur notably quickly (as quickly as 40ms after the error is performed (Rabbitt, 1966a, 1966b)), error correction cannot depend on reflective recognition and

rectification processes; it must therefore be automatic. Crucially, these automatic error corrections always increased response accuracy, and never turned a right response into a wrong one (Rabbitt, 2002, p. 1082).

Thus, automatic processes (both mental and bodily) can be sensitive to error, and able to perform quick and efficient self-correction without the intervention of slower and coarser-grained higher-order processes.[6] More specifically, the aforementioned evidence suggests that we can perform fast and intuitive error-detection and -correction processes that are independent from the effortful and cognitively effortful processing that relies on working memory. Following dual-process accounts of human cognition, we distinguish here between *automatic* or *intuitive* processes that can be performed independently from working memory, and *reflective* processes whose performance requires the use of working memory (Evans, 2010a; Nagel, 2014).[7] Here 'working memory' (Baddeley, 2007) refers to the set of higher-order cognitive capacities that allow for the mental manipulation of task-oriented representations.

Thus, there are two different ways of exerting cognitive control over our behaviour: one of them is *reflective* control—the often slow, effortful top-down control that we exert by recruiting working memory in novel or attention-demanding tasks. The other is *intuitive* or *automatic* control—the fast, rather effortless and intuitive control that we exert automatically, without the intervention of working memory.

The cases discussed above reveal that automatic processes are not (at least not necessarily) ballistic, evidence-insensitive phenomena. It thus seems too harsh to say that automatic processes are "unable to keep pace with variation in the world or with norm-world discrepancies" (Gendler, 2008b, p. 570), even if they do behave that way in some cases, or with respect to certain kinds of evidence—particularly when it is in

---

[6] For other cases of automatic error detection and correction in sensorimotor processes, see Gordon & Soechting (1995), Rabbitt (1990), and Logan & Crump (2010).

[7] Notice that this dual-process approach is significantly different from coarser dual-*system* approaches. (On the general outlook, and the difference between dual-system and dual-process accounts, see Evans (2008, 2010b); Evans & Stanovich (2013).

the form of novel propositional knowledge. We should rather agree with Brownstein and Madva in claiming that automatic processes "can be norm-sensitive in virtue of their responsiveness to affective states of disequilibrium. Responsiveness to such affective states is flexible, self-modifying, and […] a genuinely normative phenomenon" (Brownstein & Madva, 2012, p. 428). Those affective states of disequilibrium can, of course, be misguided with respect to the overall available evidence (as in the case of the trembling wise man hanging over a precipice from a perfectly safe cage), but need not be so: affective states of "felt tension" or "directed discontent" can be a part of reliable dynamic on-line error-detection and -correction systems. If we use such mechanisms to control mental action, they deserve the name 'automatic metacognition'.

3. Automatic metacognition

3.1. Control as the mark of action

Inspired by neuroscientific accounts of motor control (Jeannerod, 2006; Wolpert & Ghahramani, 2000; Wolpert, Ghahramani, & Jordan, 1995), recent theorists of action have centered their definitions of action on processes of control (Bermúdez, 2017; Hopkins, 2014; Mossel, 2005; Pacherie, 2008; Proust, 2005; Shepherd, 2014; Wu, 2016). From this viewpoint, performing an action requires not just *consciously intending* or *trying* to do something, but also doing it in a *controlled way*, exerting control over the production of the bodily or mental events.[8]

This perspective on action nicely dovetails with the data on automatic control summarized so far. In the cited examples, the subject adjusts, corrects, modifies, or

[8] Hopkins (2014) has recently appealed to the notion of control to claim that remembering is a mental action. However, we are not using the same notion of control. His notion of control refers to a type of external causation or determination of the external past events upon the remembered content ("the past must somehow control the way I now represent it […] So my representing must be directly determined by the episode: how I represent it is causally controlled by how it was, and not via the mediation of any other conscious state", p. 324). Our notion of control refers to metacognitive control and, in this sense, is internal.

simply acts in a given way while exerting an automatic form of control over performance. Even though the subject controls her behavior without a need for reflection, second-order thoughts and/or metarepresentational awareness, her action is sensitive to some normative constraints. Therefore, given the norm-sensitivity of these behaviours, the agent can be considered to be exerting *agentive automatic control* while performing them.

So, the next question that we should confront is whether there is a similar form of automatic control over mental actions, particularly over remembering.

3.2. Mental action and metacognition

Although the motor control theory works well in the case of bodily actions, we should resist the temptation of claiming that the control of mental action is carried out by the motor system, as some philosophers have suggested (Campbell, 1999; Peacocke, 2007). It is unlikely that you have to rely on motor images about your own body dispositions, as the motor systems does, in order to control your mental performance (Carruthers, 2009a; Proust, 2009). No bodily movement can fulfil a mental action's satisfaction conditions. For example, your mental attempt to retrieve a telephone number does not correspond to any bodily movement or to any physical event, because you can mentally retrieve the telephone number without moving any muscle or changing anything in the world. Therefore, the direction of fit of remembering is not world-to-mind, as in the case of bodily intentions (Searle, 1983), since it is not necessary that it should involve a change in the world (outside your mind) to be satisfied. This fact is basically explained by the nature of mental actions in contrast with bodily actions: the former aims at producing an epistemic, emotional, attentional or motivational change in the mind, whereas the latter aims at producing a change in the body and/or in the world (Kirsh & Maglio, 1994).

Thus, a first distinction between bodily action and mental action refers to their goal and the kind of control the agent must exert to perform it. Whereas in bodily action we aim at changing our body posture and/or the world, and use mainly motor

control, in mental action we aim at changing the mind and use metacognition.[9] Metacognition is the capacity to monitor and control mental processes and dispositions (Proust, 2013).

3.3. Two levels or types of metacognition

The classic view is that metacognition is thinking about thinking, i.e. forming metarepresentations, or second-order thoughts, about first-order mental states (Flavell, 1979; Nelson & Narens, 1990). Accordingly, metacognition turns the mindreading capacity towards oneself, and deploys mental concepts to produce self-ascriptions (Carruthers, 2009b, 2011). This type of metacognition has been called "theory-based metacognition" (Koriat, 2000), "high-level metacognition" (Arango-Muñoz, 2011), or "system 2 metacognition" (Proust, 2013; Shea et al., 2014) in analogy with system-2 control processes.

In contrast, recent studies have proposed that there is a leaner form of metacognition that does not require consciousness, theory of mind, or mental concepts, and operates implicitly: "Much cognitive control takes place outside system 2" (Shea et al., 2014, p. 188). Metacognition includes the mental capacity to monitor and control mental processes implicitly (Shea et al., 2014), or by means of metacognitive feelings (Arango-Muñoz, 2011, 2014; Proust, 2013).[10] This type of metacognition has been called "experience-based metacognition" (Koriat, 2000) or "system-1 metacognition" (Proust, 2013; Shea et al., 2014). [11]

---

[9] This is obviously an oversimplification: one may intend to stay still, and in this case there would not be any bodily movement involved.

[10] The implicit and the feeling-based views of the low-level metacognition are different. The former considers that there can be control in the total absence of consciousness, whereas the latter proposes that feelings are the outputs of metacognition and influence control processes. The feeling-based view is especially relevant for discussions of mental action. It is these feelings which allow us to claim that automatic error-detection and -correction processes, like the ones discussed in section 2, are at work in many situations and help the subject to detect and correct her errors: just like you are able to correct your posture when you ride your bicycle through vestibular feelings of balance, similarly you are able to correct your recollection process when you try to remember through metacognitive feelings of knowing, of forgetting, of fluency, and so on (more about this in part 4).

[11] The literature employs a distinction between 'system 1' and 'system 2' metacognition, so we use those terms here, although generally we favor a distinction in terms of process types, consistently with our preference for dual-*process* over dual-*system* accounts of cognition (Evans & Stanovich 2013).

"System-2 metacognition" is widely accepted, so we will only briefly discuss evidence for "system-1 metacognition", which comes mainly from three domains: 1) comparative psychology has shown that animals devoid of mindreading capacity and mental concepts pass metacognitive tasks: Rhesus monkeys and bottlenose dolphins are able to monitor and control their perceptual and memory capacities (Hampton, 2001; Smith, 2009; Smith, Beran, Redford, & Washburn, 2006; Smith & Washburn, 2003). 2) Psychologists have found that subjects very often monitor their mental processes based on metacognitive feelings and heuristics, and not necessarily on theoretical information (Koriat, 2000). A growing literature shows the important role of metacognitive feelings in the control of mental processes (Fernández Cruz et al., 2016; Koriat, 2000; Schwartz & Metcalfe, 2010). Finally, 3) there seem to be differences in the neural activity related to mindreading (system-2 metacognition) and system-1 metacognition (Proust, 2012; Schnyer et al., 2004). According to some studies, mindreading is correlated with neural activity in the right temporal-parietal junction, the prefrontal antero-medial cortex, and anterior temporal cortex (Del Cul, Dehaene, Reyes, Bravo, & Slachevsky, 2009; Perner & Aichorn, 2008); whereas system-1 metacognition is correlated with neural activity in the dorsolateral prefrontal cortex, the ventro-medial prefrontal cortex, and the anterior cingulate cortex (Schnyer et al., 2004).

As argued before, when we exert agentive control through metacognition we perform mental actions. So, if remembering is an action, we should exert control over our episodic reconstruction processes by means of one of these two kinds of metacognition. One can control one's remembering either by system-2 metacognition (i.e., thinking about what one is remembering and making some metacognitive inferences about it), or by system-1 metacognition (i.e. monitoring and controlling the process of remembering by means of metacognitive feelings). However, since system-2 metacognition is cognitively very demanding and costly, most of the metacognitive control is feeling-based. That is the reason why we will focus on system-1. In what follows, we argue that system-1, feeling-based metacognition allows us to exert agentive control over the content-production aspect of episodic memory processes.

4. Feeling-based metacognition and episodic memory

4.1. Episodic memory as the imaginative reconstruction of the personal past

The reconstructive theory of memory claims that memory is not the passive retrieval of stored representations, but rather the active reconstruction of a representation of a past episode: "Remembering is a matter of imagining or simulating the past" (Michaelian, 2016, p. 60). However, it is also worth noting that reconstruction takes place not only during retrieval, but also during the encoding stage of the process, in which memory selects what to retain and how to retain it, often encoding only a gist of the remembered event (Schacter & Addis, 2007).

If remembering is the action of imaginative reconstruction of a past experience, this opens the door for many mnemonic errors such as false recognition (Roediger & McDermott, 1995), boundary extension (Intraub, Bender, & Mangels, 1992), the superportrait phenomenon (Rhodes, 1996), and confabulation (Michaelian, 2011). Misremembering occurs when a reliable mnemonic system produces a false or inaccurate representation of the past. Reconstruction theorists highlight that, given the constructive character of memory, misremembering is a systematic and ordinary occurrence in our daily lives (De Brigard, 2014a).

Given the imperfect character of memory outputs, some kind of control is needed to ensure the reliability of memory. This is not the preparatory, stage-setting control of mental action discussed above (trying to bring it about that I remember *X*); rather, what is required here is control over the automatic imaginative content-production processes and its reliability (trying to remember correctly). Metacognition, the control of mental processes and dispositions, achieves this largely by means of metacognitive feelings.

4.2. How metacognitive feelings guide episodic reconstruction

As we said earlier, memory has a rich phenomenology: recollection is often not an immediate mental event, but a lengthy process, and while it is taking place we experience not only that we can bring the relevant mnemonic content to mind; we also feel like remembering is sometimes easier, sometimes harder; we feel that a given memory is more or less accurate; we feel that we will be able to remember something if we try harder (e.g., in the tip-of-the-tongue phenomenon), or we simply know that

we will not be able to remember, no matter how hard we try; sometimes we feel that we are forgetting something we should remember; and so on. All these phenomena are metacognitive experiences that convey some information about the way the episodic reconstruction process is going, whether it is running smoothly or it has found some obstacles, and how serious these obstacles are. As mentioned above, these phenomena suggest that there is more to episodic memory control than simply holding an intention in mind and waiting for the right content to come. On this respect, Souchay et al. remark that "presumably on-line feelings and thoughts generated during retrieval by the object level are monitored by the meta-level, leading to the implementation of mnemonic strategies, termination of search, and so on" (2013, p. 1). These phenomena suggest that we not only have control over the global act of remembering, but also that, with the guidance of metacognitive feelings, agents have some level of control over the reconstruction process. Thus, metacognitive feelings enable the agent's strategic involvement in the mnemonic process, bolstering the reliability of the mental content produced.

The feeling of knowing (FOK) is one of the most interesting and puzzling experiences related to memory. Sometimes, when you are faced with a memory problem, even before trying to retrieve the solution, you already feel that you know it, that you will be able to reconstruct the memory. In the domain of semantic memory, it has been shown that FOKs play a central role in deciding whether to remember a piece of information or try another strategy to retrieve the information (Arango-Muñoz, 2013; Paynter, Reder, & Kieffaber, 2009; Reder, 1987). Although FOK has been less studied in the domain of episodic memory,[12] it seems likely that this experience provides the subjects with a sense of whether they would be able to reconstruct a piece of information or not. When asked whether she will be able to remember the events that took place at her school graduation, the feeling of knowing

[12] Some researchers consider that studying and recalling word lists or sentences is a means to study episodic memory and not only semantic. Although we agree that studying and recalling word lists may indirectly inform our knowledge of episodic memory, it is not the most straightforward way to study this type of memory.

will motivate the subject to give a positive answer and to start a reconstruction attempt.

Subjects also feel that remembering is sometimes easier or harder. The feeling of ease has been shown to play a fundamental role in the monitoring and control of memory retrieval. According to Koriat's FOK accessibility model (1993, 2000), the accessibility of partial or contextual information relevant to the memory target gives rise to this feeling. E.g. immediate or quick access to a song or anthem's first lines gives rise to the feeling of knowing it. Think about what happens when you want to tell people a joke you heard earlier: you mentally rehearse the first few lines and the punchline, and, if they come to mind easily, this feeling may motivate you to start telling the joke, with the (often confirmed) certainty that you will be able to reconstruct the rest. But if these traces do not come to mind quickly, you delay the start of the joke until you successfully retrieve enough of them. (This is a skill that must be learned, which explains why children way too often ruin jokes by starting to tell them, and only later realizing they have forgotten the punchline.) Conversely, a person may have a feeling of uncertainty, of not knowing or forgetting, which may motivate her to give up on the mnemonic reconstruction rapidly.

We also feel that a given memory trace is more or less accurate, which leads us to structure the recollection process in different ways. Some remembered information feels right, and some feels wrong. It has been found that the fluency with which a memory is reconstructed is a key determinant of the feeling of rightness. "Fluency" refers to the ease with which an information is reconstructed, and thus has been mainly measured through reaction time, i.e. the speed of the mnemonic reconstruction. A memory that is reconstructed fluently (i.e. quickly) feels right, whereas a reconstruction that isn't fluent feels dubious (Benjamin, Bjork, & Schwartz, 1998; Kelley & Lindsay, 1993; Whittlesea & Leboe, 2003). When episodic retrieval is accompanied by a feeling of wrongness or doubt, this tends to motivate the agent to revise the outcome until the outcome feels more accurate (Gallo & Lampinen, 2015).

The feeling of forgetting also plays an important role in remembering. When we try to mentally reconstruct lists of items, scenes, or to-do lists, we often have a feeling that we have forgotten something (Halamish, McGillivray, & Castel, 2011).

This feeling motivates the subject to revise the mnemonic reconstruction, to seek confirmatory cues and traces, and to check if something is missing; this feeling also casts doubt on the memory's integrity.

All these metacognitive feelings are of vital importance in the agent's production and evaluation of the mnemonic reconstruction, and her further decision of whether to endorse the reconstructed information (Michaelian, 2012). Our claim is that when a subject is remembering guided by these metacognitive feelings, she is actually exerting *agentive automatic control* over her mental mnemonic processes, since said feelings motivate the subject to resolve felt tensions and guide her mnemonic reconstruction toward reliability. Thus, given the centrality of agentive control for action, when the agent controls her mental processes of episodic reconstruction by means of metacognitive feelings, remembering counts as a mental action. Further, in feeling-based remembering the metacognitive and the reconstructive processes jointly constitute the action, since it is through metacognition that the agent manages to structure the automatic reconstructive processes toward the production of a reliable memory that she can then endorse.

5. Conclusion

The main reason for some philosophers to consider that memory is just a passive information retention and retrieval capacity is its automaticity. The apparent "ballistic" nature of the production of mental content seems to leave no room for agency and control. In opposition to this view, we have defended an active account of memory according to which remembering can count as a mental action. Our claim differs from a "simple-connection" view because we hold that agents are able not only to control the top-down initiation of remembering, but also to structure the content-production process, monitoring and guiding it by means of intuitive metacognitive feelings.

Memory, as an imaginative reconstruction of past experience, frequently requires some kind of metacognitive control to ensure its reliability. Drawing from recent work on the normativity of automaticity and automatic control, we distinguish two kinds of metacognitive control: top-down, reflective control, and automatic, intuitive, feeling-based control. We propose that whenever agents control their mental

remembering processes by means of intuitive or feeling-based metacognition, remembering is an action.

Before ending, it is worth making some clarifications. First, our view does not entail that metacognitive feelings are perfect guides to reliable memories. We often misattribute reliability levels to our memories, and some people are surely better than others at producing reliable memories; but to the extent that we can intentionally improve the reliability of our recollections through purely internal mental processes, it is largely thanks to our metacognitive feelings. Second, we do not want to exclude high-level metacognition as another source of mnemonic agency: there are certainly some high-level inferential processes at play in some of our remembering routines (like the ones present in source monitoring). However, given our tendency to minimize executive control processes, most everyday mnemonic control processes surely are system-1 processes (Shea et al., 2014, p. 188), so low-level metacognition seems to be more common than high-level metacognition, covering a wider scope of cases. Third, we do not claim that all memory-like episodes are mental actions: our claim is that those monitored and controlled by agents via feeling-based metacognition are actions; certainly higher-order metacognition can also guarantee that some others are actions too. But we also want to leave room for mere episodic mental happenings or events that are not agentive, like the sudden involuntary rehearsal of the song I heard yesterday.

Acknowledgements

The authors would like to express their gratitude to Felipe de Brigard, Kirk Michaelian, Markus Wernig, Denis Perrin, Wayne Wu, Michael Brent, Flavia Felletti, Santiago Amaya, Sam Murray, and two anonymous reviewers for their thoughtful, often very detailed, and tremendously helpful comments and suggestions.

References

Arango-Muñoz, S. (2011). Two levels of metacognition. *Philosophia*, (39), 71–82. https://doi.org/10.1007/s11406-010-9279-0

Arango-Muñoz, S. (2013). Scaffolded Memory and Metacognitive Feelings. *Review of Philosophy and Psychology*, *4*(1), 135–152. https://doi.org/10.1007/s13164-012-0124-1

Arango-Muñoz, S. (2014). The nature of epistemic feelings. *Philosophical Psychology*, *27*(2), 1–19. https://doi.org/10.1080/09515089.2012.732002

Baddeley, A. (2007). *Working memory, thought, and action*. Oxford: Oxford University Press.

Bargh, J. A. (1992). The ecology of automaticity: Toward establishing the conditions needed to produce automatic processing effects. *The American Journal of Psychology*, 181–199.

Bargh, J. A. (1997). The automaticity of everyday life. In R. S. Wyer (Ed.), *Advances in Social Cognition* (pp. 1–62). Lawrence Erlbaum Associates.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, *71*(2), 230–.

Benjamin, A. S., Bjork, R. A., & Schwartz, B. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, *127*, 55–68.

Bermúdez, J. P. (2017). Do we reflect while performing skillful actions? Automaticity, control, and the perils of distraction. Philosophical Psychology, online

first, 1–29. *Philosophical Psychology*. https://doi.org/10.1080/09515089.2017.1325457

Bresner, D., & Stolz, J. A. (1999). What kind of attention modulates the Stroop effect? *Psychonomic Bulletin & Review*, *6*(1), 99–104.

Brownstein, M., & Madva, A. (2012). The normativity of automaticity. *Mind and Language*, *27*(4), 410–434.

Campbell, J. (1999). Schizophrenia, the space of reasons and thinking as a motor process. *The Monist*, *82*(4), 609–25.

Carruthers, P. (2009a). Action-awareness and the active mind. *Philosophical Papers*, *38*(2), 133–156. https://doi.org/10.1080/05568640903146443

Carruthers, P. (2009b). How we know our own minds: The relationship between mindreading and metacognition. *Behavioral and Brain Sciences*, *32*(2), 121. https://doi.org/10.1017/S0140525X09000545

Carruthers, P. (2011). *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford: Oxford University Press.

Clark, A. (2015). What the "extended me" knows. *Synthese*. https://doi.org/10.1007/s11229-015-0719-z

De Brigard, F. (2014a). Is memory for remembering? Recollection as a form of episodic hypothetical thinking. *Synthese*, *19*(2), 155–185. https://doi.org/10.1007/s11229-013-0247-7

De Brigard, F. (2014b). The nature of memory traces. *Philosophy Compass*, *9*(6), 402–414.

Del Cul, A., Dehaene, S., Reyes, P., Bravo, E., & Slachevsky, A. (2009). Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain*, *132*(9), 2531–2540.

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255–278.

Evans, J. S. B. T. (2010a). Intuition and reasoning: A dual-process perspective. *Psychological Inquiry*, *21*(4), 313–326.

Evans, J. S. B. T. (2010b). *Thinking Twice: Two minds in One Brain*. Oxford: Oxford University Press.

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223–241.

Fernández Cruz, A. L., Arango-Muñoz, S., & Volz, K. (2016). Oops, scratch that! Monitoring one's own errors during mental calculation. *Cognition*, *146*, 110–120. https://doi.org/10.1016/j.cognition.2015.09.005

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, *34*(10), 906–911. https://doi.org/10.1037//0003-066x.34.10.906

Francolini, C. M., & Egeth, H. (1980). On the nonautomaticity of `automatic' activation: Evidence of selective seeing. *Perception and Psychophysics*, *27*, 331–342.

Gallo, D. A., & Lampinen, J. M. (2015). Three Pillars of False Memory Prevention:: Orientation, Evaluation and Corroboration. In J. Dunlosky & S. K. Tauber

(Eds.), *The Oxford Handbook of Metamemory*. Retrieved from https://doi.org/10.1093/oxfordhb/9780199336746.013.11

Gendler, T. S. (2008a). Alief and belief. *Journal of Philosophy*, *105*(10), 634–663.

Gendler, T. S. (2008b). Alief in action (and reaction). *Mind and Language*, *25*(5), 552–585.

Gordon, A. M., & Soechting, J. F. (1995). Use of tactile afferent information in sequential finger movements. *Experimental Brain Research*, *107*(2), 281–292.

Halamish, V., McGillivray, S., & Castel, A. D. (2011). Monitoring One's Own Forgetting in Younger and Older Adults. *Psychology and Aging*. https://doi.org/10.1037/a0022852

Hampton, R. R. (2001). Rhesus monkeys know when they remember. *Proceedings of the National Academy of Sciences U.S.A*, *98*, 5359–5362.

Hopkins, R. (2014). Episodic Memory as Representing the Past to Oneself. *Review of Philosophy and Psychology*, *5*(3), 313–331. https://doi.org/10.1007/s13164-014-0184-5

Intraub, H., Bender, R. S., & Mangels, J. A. (1992). Looking at pictures but remembering scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 180–191.

Jeannerod, M. (2006). *Motor cognition*. Oxford: Oxford University Press.

Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar: Straus and Giroux. Retrieved from http://doi.org/10.1007/s13398-014-0173-7.2

Kahneman, D., & Henik, A. (1981). Perceptual organization and attention. In M. Kubovy & J. Pomerantz (Eds.), *Perceptual Organization and Attention*. Erlbaum.

Kelley, C. M., & Lindsay, S. D. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge. *Journal of Memory and Language*, *32*, 1–24.

Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science*, *18*, 513–549.

Koch, C., & Crick, F. (2001). On the zombie within. *Nature*, *41*(6840), 893–893.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, *100*(4), 609–639.

Koriat, A. (2000). The feeling of knowing: Some metatheoretical implications for consciousness and control. *Consciousness and Cognition*, *9*(2), 149–171.

LaBerge, D., & Samuels, S. J. (1974). LaBerge, D. & Samuels, S. J. (1974). "Toward a theory of automatic information processing in reading." Cognitive psychology 6:2, 293--323. *Cognitive Psychology*, *6*(2), 293–323.

Logan, G. D., & Cowan, W. B. (1984). Logan, G. D. & Cowan, W. B. (1984). "On the ability to inhibit thought and action: A theory of an act of control." Psychological review 91:3, 295. *Psychological Review*, *91*(3), 295–?????

Logan, G. D., & Crump, M. J. C. (2010). Cognitive Illusions of Authorship Reveal Hierarchical Error Detection in Skilled Typists. *Science*, *330*, 683–686. https://doi.org/10.1126/science.1190483

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological Bulletin*, *109*(2), 163–203.

Mele, A. R. (1997). Agency and mental action. *Nous*, *31*, 231–249.

Mele, A. R. (2009). Mental action: a case study. In *Mental actions* (O'Brien L. & Soteriou M., pp. 17–37). Oxford: Oxford University Press.

Michaelian, K. (2011). Generative memory. *Philosophical Psychology*, *24*(3), 323–342.

Michaelian, K. (2012). Metacognition and endorsement. *Mind and Language*, *27*(3), 284–307.

Michaelian, K. (2016). *Mental Time Travel: Episodic Memory and our Knowledge of the Personal Past*. Cambridge, Massachusetts: The MIT Press.

Moors, A., & De Houwer, J. (2006). Automaticity: a theoretical and conceptual analysis. *Psychological Bulletin*, *132*(2), 297–?????

Mossel, B. (2005). Action, control and sensation of acting. *Philosophical Studies*, *124*, 129–180.

Nagel, J. (2014). Intuition, reflection, and the command of knowledge. *Proceedings of the Aristotelian Society*, *88*, 217–239.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *The Psychology of Learning and Motivation*, *26*, 125–173.

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and Self-Regulation* (pp. 1–18). Springer.

Norman, E., Price, M. C., & Duff, S. C. (2010). Fringe consciousness: A useful framework for clarifying the nature of experience-based metacognitive feelings. In A. Efklides & P. Misailidi (Eds.), *Trends and prospects in metacognition research* (pp. 63–80). Springer.

Pacherie, E. (2008). The phenomenology of action: A conceptual framework. *Cognition*, *107*, 179–217. https://doi.org/doi:10.1016/j.cognition.2007.09.003

Paynter, C. A., Reder, L. M., & Kieffaber, P. D. (2009). Knowing we know before we know: ERP correlates of initial feeling-of-knowing. *Neuropsychologia*, *47*(3), 796–803. https://doi.org/10.1016/j.neuropsychologia.2008.12.009

Peacocke, C. (2007). Mental Action and Self-Awareness (I). In *Contemporary Debates in the Philosophy of Mind* (Brien McLaughlin & Jonathan Cohen). Blackwell.

Perner, J., & Aichorn. (2008). Theory of mind, language and the temporo-parietal junction mystery. *Trends in Cognitive Sciences*, *12*(4), 123–126.

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control". In Information processing and cognition: The Loyola symposium. In R. Solso (Ed.), *Information processing and cognition: the Loyola Symposium*. Erlbaum.

Proust, J. (2005). *La nature de la volonté*. Paris: Folio-Gallimard.

Proust, J. (2009). Is there a sense of agency of thought? In L. O'Brien & M. Soteriou (Eds.), *Mental actions and agency* (pp. 253–279). Oxford: Oxford University Press.

Proust, J. (2012). Metacognition and mindreading: one or two functions? In M. Beran, J. Brandl, J. Perner, & J. Proust (Eds.), *The foundations of metacognition* (pp. 234–250). Oxford: Oxford University Press.

Proust, J. (2013). *The Philosophy of Metacognition: Mental Agency and Self-Awareness*. Oxford: Oxford University Press.

Rabbitt, P. (1966a). Error correction time without external error signals. *Nature*, *212*, 438–XXX.

Rabbitt, P. (1966b). Errors and error correction in choice-response tasks. *Journal of Experimental Psychology*, *71*(2), 264–XXX.

Rabbitt, P. (1990). Age, IQ and awareness of errors. *Ergonomics*, *33*(10), 1291–1305.

Rabbitt, P. (2002). Consciousness is slower than you think. *The Quarterly Journal of Experimental Psychology*, *55*(4), 1081–1092.

Reder, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology*, *19*(1), 90–138. https://doi.org/10.1016/0010-0285(87)90005-3

Rhodes, G. (1996). *Superportraits: Caricatures and recognition*. Hove: Psychology Press.

Rietveld, E. (2008). Situated normativity: The normative aspect of embodied cognition in unreflective action. *Mind*, *117*(468), 973–1001.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 803–814.

Schacter, D. L., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory: remembering the past and imagining the future. *Philosophical Transactions of the Royal Society B*, *262*, 773–786. https://doi.org/10.1098/rstb.2007.2087

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*(1), 1–66.

Schnyer, D. M., Verfaellie, M., Alexander, M. P., LaFleche, G., Nichols, L., & Kaszniak, A. W. (2004). A role for right medial prefrontal cortex in accurate feeling-of-knowing judgments: evidence from patients with lesions to frontal cortex. *Neuropsychologia*, *42*, 957–966.

Schwartz, B., & Metcalfe, J. (2010). Tip-of-the-tongue (TOT) states: Retrieval, behavior, and experience. *Memory and Cognition*, *39*, 737–749.

Searle, J. (1983). *Intentionality*. Cambridge: Cambridge University Press.

Shea, N., Boldt, A., Bang, D., Yeung, N., Heyes, C., & Frith, C. D. (2014). Supra-personal cognitive control and metacognition. *Trends in Cognitive Sciences*, *18*(4), 186–193. https://doi.org/http://dx.doi.org/10.1016/j.tics.2014.01.006

Shepherd, J. (2014). The contours of control. *Philosophical Studies*, *170*(3), 395–411. https://doi.org/10.1007/s11098-013-0236-1

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*(2), 127–194.

Smith, J. D. (2009). The study of animal metacognition. *Trends in Cognitive Sciences*, *13*(9), 389–396. https://doi.org/10.1016/j.tics.2009.06.009

Smith, J. D., Beran, M. J., Redford, J. S., & Washburn, D. A. (2006). Dissociating uncertainty responses and reinforcement signals in the comparative study of uncertainty monitoring. *Journal of Experimental Psychology: General*, *135*, 282–297.

Smith, J. D., & Washburn, D. A. (2003). The comparative psychology of uncertainty monitoring and metacognition. *Behavioral and Brain Sciences*, *26*, 317–373.

Souchay, C., Guillery-Girard, B., Pauly-Takacs, K., Wojcik, D. Z., & Eustache, F. (2013). Subjective experience of episodic memory and metacognition: a neurodevelopmental approach. *Frontiers in Behavioral Neuroscience*, *7*, 1–16. https://doi.org/10.3389/fnbeh.2013.00212

Stanovich, K. E. (2004). *The Robot's Rebellion: Finding Meaning in the Age of Darwin. University of Chicago Press.* University of Chicago Press. Retrieved from http://doi.org/10.1007/s13398-014-0173-7.2

Strawson, G. (2003). Mental Ballistics or the Involuntariness of Spontaneity. *Proceedings of the Aristotelian Society, New Series*, *103*, 227–256.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Walsh, M. M., & Anderson, J. R. (2009). The strategic nature of changing your mind. *Cognitive Psychology*, *58*, 416–440.

Whittlesea, B. W. A., & Leboe, J. P. (2003). Two fluency heuristics (and how to tell them apart). *Journal of Memory and Language*, *49*, 62–79.

Wolpert, D. M., & Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nature*, *3*, 1212–1217.

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, *269*, 1880–1882.

Wu, W. (2013). Mental Action and the Threat of Automaticity. In A. Clark, J. Kiverstein, & T. Vierkant (Eds.), *Decomposing the Will* (pp. 244–261). Oxford: Oxford University Press.

Wu, W. (2016). Experts and Deviants: The Story of Agentive Control. *Philosophy and Phenomenological Research*, *93*(1), 101–126. https://doi.org/10.1111/phpr.12170