

Thomas Metzinger

## *M-Autonomy*

**Abstract:** *What we traditionally call ‘conscious thought’ actually is a subpersonal process, and only rarely a form of mental action. The paradigmatic, standard form of conscious thought is non-agentive, because it lacks veto-control and involves an unnoticed loss of epistemic agency and goal-directed causal self-determination at the level of mental content. Conceptually, it must be described as an unintentional form of inner behaviour. Empirical research shows that we are not mentally autonomous subjects for about two thirds of our conscious lifetime, because while conscious cognition is unfolding, it often cannot be inhibited, suspended, or terminated. The instantiation of a stable first-person perspective as well as of certain necessary conditions of personhood turn out to be rare, graded, and dynamically variable properties of human beings. I argue that individual representational events only become part of a personal-level process by being functionally integrated into a specific form of transparent conscious self-representation, the ‘epistemic agent model’ (EAM). The EAM may be the true origin of our consciously experienced first-person perspective.*

### **1. M-Autonomy**

The two main claims of this contribution are, first, that for roughly two thirds of their conscious lives human beings are not mentally autonomous subjects, and, second, that what we traditionally call ‘conscious thought’ primarily and predominantly is a subpersonal process. The argument is partly based on recent empirical research demonstrating the ubiquitous occurrence of ‘mind-wandering’, or spontaneous, task-unrelated thought. Examples of mind-wandering are daydreams, automatic planning, the sudden occurrence of unbidden

Correspondence:  
Email: metzinge@uni-mainz.de

memories, or depressive rumination.<sup>1</sup> Methods like externally cued experience sampling show how unnoticed attentional lapses leading to uncontrolled mental activity of this kind are much more frequent than most of us intuitively think. Mind-wandering is interesting for philosophy of mind because its phenomenology as well as new empirical data bear direct relevance on our theoretical notion of ‘mental autonomy’ (M-autonomy). As it were, mind-wandering is the opposite of M-autonomy, because it involves a loss of self-control at the level of conscious thought. My epistemic goal in this paper is to find out how a more careful look at the phenomenology and novel empirical data can help to improve our conceptual understanding of what it means to be a mentally autonomous subject.

Mental autonomy includes the capacity to impose rules on one’s own mental behaviour, to explicitly select goals for mental action, the ability for rational guidance and, most importantly, for the intentional inhibition, suspension, or termination of an ongoing mental process. M-autonomy is a functional property,<sup>2</sup> which any given self-conscious system can either possess or lack. Its instantiation goes along with new epistemic abilities, a specific phenomenological profile, and the appearance of a new layer of representational content in the phenomenal self-model (Metzinger, 2003a). In humans, first insights into its neuronal realization are now beginning to emerge. From a philosophical perspective, this functional property is interesting for a whole range of different reasons. One of them is that it is directly

- 
- <sup>1</sup> This paper is based on an earlier and more comprehensive publication of mine, which offers a first philosophical perspective on the recent surge of scientific work related to the phenomenon of ‘mind-wandering’ (Metzinger, 2013a). It aims at further developing only a few of its central ideas and leaves out as much empirical detail as possible, including my own empirical hypothesis that mind-wandering can be characterized by unnoticed switches in what I have called the phenomenal ‘unit of identification’ (UI) and an experimentally detectable ‘self-representational blink’ (SRB). In terms of recent references on the topic since my earlier publication, I recommend Smallwood and Schooler (2015) for a recent empirical review; Carruthers (2015, Chapter 6.5), Dorsch (2014), Irving (2015), and Pliushch and Metzinger (2015) are philosophical discussions. I am also extremely grateful to two anonymous reviewers, who have both offered very helpful, constructive, and substantial criticism, as well as to Carsten Korth and Wanja Wiese for additional comments.
- <sup>2</sup> Functional properties are abstract properties referring to the *causal role* of a state (the set of its causal relations to input, output, and other internal states), without implying anything about the properties of its physical realization. Just like states described in a Turing machine table or computer software, they are multi-realizable. For example, as M-autonomy is a functional property, it could in principle also be implemented in a machine.

relevant to both our traditional notions of a 'first-person perspective' and of 'personhood'. If one cannot control the focus of one's attention, then one cannot sustain a stable first-person perspective, and for as long as one cannot control one's own thought one cannot count as a rational individual.

This paper is composed of three parts. First, I will briefly introduce the concept of 'M-autonomy'. Part 2 will connect this new idea with the two notions of 'possessing a first-person perspective' and of 'personhood', by enriching the functionalist concept with a dynamic, representationalist account: M-autonomy consists in the possession of an 'epistemic agent model' (EAM). Here, one central point is that the transition from subpersonal to personal-level cognition is enabled by a specific form of conscious self-representation, namely, a global model of the cognitive system as an entity that *actively* constructs, sustains, and controls knowledge relations to the world and itself. In Part 3, I will show that for the largest part of our conscious lives we are not mentally autonomous cognitive systems in this sense and conclude that what we traditionally call 'conscious thought' actually is a sub-personal process.

Let us begin by pointing out how biological systems produce different kinds of observable output, which can in turn be characterized by different degrees of autonomy and self-control. For the purposes of this paper, let us say that there are actions and behaviours. Both kinds of output are conceptually individuated by their satisfaction conditions; they are directed at goal states. However, for actions, conscious goal-representation plays a central causal role, actions can be terminated, suspended, intentionally inhibited, and they exhibit a distinct phenomenological profile involving subjective qualities like agency, a sense of effort, goal-directedness, global self-control, and ownership. Behaviours, on the other hand, are purposeful, but possess no explicit form of conscious goal-representation. They are functionally characterized by automaticity, decreased context-sensitivity, and low self-control, we may not even notice their initiation, but they can be faster than actions. While their phenomenological profile can at times be completely absent, behaviours typically involve the subjective experience of ownership without agency, whereas the introspective availability of goal-directedness varies and there frequently is a complete lack of meta-awareness.

There are not only bodily actions, but also mental actions. Deliberately focusing one's attention on a perceptual object or consciously drawing a logical conclusion are examples of mental actions.

Just like physical actions, mental actions possess satisfaction conditions (i.e. they are directed at a goal state). Although they mostly lack overt behavioural correlates, they can be intentionally inhibited, suspended, or terminated, just like bodily actions can. In addition, they are interestingly characterized by their temporally extended phenomenology of ownership, goal-directedness, a subjective sense of effort, and the concomitant conscious experience of agency and *mental* self-control.

Let me distinguish the two most important types of mental action:

- **Attentional agency (AA)**, the ability to control one's focus of attention.
- **Cognitive agency (CA)**, the ability to control goal/task-related, deliberate thought.

AA and CA are not only functional properties that are gradually acquired in childhood, can be lost in old age or due to brain lesions, and whose incidence, variance, robustness, etc. can be scientifically investigated. They also have a subjective side: attentional agency (Metzinger, 2003a, 6.4.3; 2006, Section 4) is also a phenomenal property, as is the case for pain or the subjective quality of 'blueness' in a visual colour experience (Metzinger, 1995). AA is the conscious experience of actually initiating a shift of attention, of controlling and fixing its focus on a certain aspect of reality. AA involves a sense of effort, and it is the phenomenal signature of our functional ability to actively influence what we will come to know, and what, for now, we will ignore. Consciously experienced AA is theoretically important, because it is probably the earliest and simplest form of experiencing oneself as a knowing self, as an epistemic agent. To consciously enjoy AA means that you (the cognitive system as a whole) currently identify with the content of a particular self-representation, an 'epistemic agent model' (EAM; see Section 2 and Metzinger, 2013a,b) currently active in your brain. AA is fully transparent:<sup>3</sup> the

---

<sup>3</sup> 'Transparency' is a property of conscious representations, namely, that they are not experienced *as* representations. Therefore, the subject of experience has the feeling of being in direct and immediate contact with their content. Transparent conscious representations create the phenomenology of naïve realism. An opaque phenomenal representation is one that is experienced *as* a representation, for example in pseudo-hallucinations or lucid dreams. Importantly, a transparent self-model creates the phenomenology of identification (Section 3; Metzinger, 2003a; 2008). There exists a graded spectrum between transparency and opacity, determining the variable

content of your conscious experience is not one of self-representation or of an ongoing process of self-modelling, of depicting yourself as a causal agent in certain shifts of ‘zoom factor’, ‘resolving power’, or ‘resource allocation’, and so on. Rather, you directly experience *yourself* as, for example, actively selecting a new object for attention. During mind-wandering episodes we do not have AA, although these episodes can of course be *about* having been an attentional agent in the past, or *about* planning to control one’s attention in the future. Other examples of situations in which this property is selectively missing are non-lucid dreaming and NREM-sleep mentation (Metzinger, 2013b; Windt, 2015), but also infancy, dementia, or severe intoxication syndromes.

An analogous point can be made for CA. Conceptually, cognitive agency is not only a complex set of functional abilities, like the capacity of mental calculation, consciously drawing logical conclusions, engaging in rational, symbolic thought, and so on. Again, there is a distinct phenomenology of currently being a cognitive *agent*, which can lead to experiential self-reports like ‘I am a thinking self in the act of grasping a concept’, ‘I have just actively arrived at a specific conclusion’, etc. What AA and CA have in common is that in both cases we consciously represent ourselves as epistemic agents: according to subjective experience, we are entities that actively construct and search for new epistemic relations to the world and ourselves.

There are, however, not only mental actions, but also mental behaviours. ‘Mind-wandering’, or spontaneous, task-unrelated thought, is a paradigm example of unintentional mental behaviour. It may often be purposeful, but exhibits no conscious goal-representation, no overt behavioural correlates, it is characterized by an unnoticed loss of mental self-control and high degrees of automaticity, plus a lack of sensitivity to the situational context, while the phenomenological profile is characterized by ownership without agency, variable or absent introspective availability of goal-directedness, and frequently by a complete lack of meta-awareness (Schooler *et al.*, 2011). Empirically, it is plausible to assume that unconscious mind-wandering, instantiating no phenomenal properties whatsoever, exists as well (Horovitz *et al.*, 2009; Pliushch and Metzinger, 2015; Samann *et al.*, 2011; Vanhaudenhuyse *et al.*, 2010).

---

phenomenology of ‘mind-independence’ or ‘realness’. Unconscious representations are neither transparent nor opaque. See Metzinger (2003b) for a concise introduction.

What we can consciously access as daydreaming, inner thoughts, fantasies, unbidden memories and feelings may rather be just the tip of the iceberg, a small partition of a much larger state space in which the continuous cognitive dynamics unfolds. Conscious mind-wandering would then be characterized by a higher degree of coherence, but still emerge out of a larger unconscious background of activity. Mind-wandering and nocturnal dreaming (*cf.* Metzinger, 2013a,b; Fox *et al.*, 2013; Wamsley, 2013; Windt and Metzinger, 2007; Windt, 2015) are both interesting to philosophers of mind, because both involve sudden shifts in mechanisms of self-identification, rationality deficits, and a cyclically recurring decrease in mental autonomy that is not self-initiated and frequently unnoticed.

Some mental activities are not autonomously controllable, because one centrally important defining characteristic does not hold: they cannot be inhibited, suspended, or terminated. Let us call these activities ‘unintentional mental behaviours’. Mind-wandering can therefore be conceptualized as a form of unintentional behaviour, as an involuntary form of mental activity. Of course, the fact that a given behaviour, be it mental or bodily, is unintentional in no way implies that this behaviour is unintelligent or even maladaptive. For example, low-level, saliency-driven shifts in attentional focus are unintentional mental behaviours, and not inner actions. In standard situations, they cannot be inhibited. They are initiated by unconscious mechanisms, but may well result in a stable, perceptually coupled first-person perspective as their final stage. Stimulus-independent, task-independent thought, however, normally begins as a form of uncontrolled mental behaviour, a breakdown of consciously guided epistemic auto-regulation, which is the active control of one’s own epistemic states at the level of high-level cognition. Just like an automatic, saliency-driven shift in the focus of attention, it may be caused by unconscious factors like introspectively inaccessible goal representations that drive the high-level phenomenology of mind-wandering (Klinger, 2013), for example by representations of postponed goal-states which have been environmentally cued by goal-related stimuli under high cognitive load (Cohen, 2013; McVay and Kane, 2013). Both low-level attention and uncontrolled, automatic thinking will frequently count as intelligent, an adaptive type of inner behaviour. But as long as it is going on, we seem to lack the ability to terminate or suspend it — we are fully immersed in an inner narrative and cannot deliberately ‘snap out of it’ (see below and note #9). Perhaps the most relevant and hitherto

neglected phenomenological constraint for a theory of mental autonomy is that, subjectively, we do not notice this fact.

But what exactly is autonomy? Very generally speaking, autonomy would be the capacity for rational self-control, whereas the term ‘mental autonomy’ refers to the specific ability to control one’s own mental functions, like attention, episodic memory, planning, concept formation, rational deliberation, or decision making, etc. Let us begin by looking at the contrast class of our target phenomenon, at cases where some of these functions would selectively operate *without* the decisive ability to wilfully terminate or suspend them. How can one better describe the missing element? As it turns out, the contrast class is very large. A second highly relevant fact that has been almost completely overlooked by philosophers, or so I will claim in Section 3, is that a recurring *loss* of mental autonomy is one major characteristic of our cognitive phenomenology,<sup>4</sup> and that both research on dreaming and mind-wandering have already developed important research tools to investigate this hitherto neglected aspect further (like external probing, or systematic questions after sleep lab awakenings; cf. Smallwood, 2013; Windt, 2015). However, in this case, empirical and conceptual questions are so deeply intertwined that we need a stronger form of cooperation between the disciplines. Therefore, what is now needed is a first set of conceptual instruments that opens the field for fruitful interdisciplinary collaboration.

One way of providing a richer conceptual analysis of what a loss of mental autonomy actually amounts to is by describing it as losing the ability for *second-order mental action*. This ability can be decomposed into the following capacities:

- The imposing of rules on one’s own mental behaviour;
- explicit goal-selection, goal-commitment, goal-permanence;
- satisfaction of rationality constraints or rational guidance;
- intentional inhibition, suspension, or termination of an ongoing process.

---

<sup>4</sup> ‘Cognitive phenomenology’ is a new subfield of research in philosophy of mind that focuses on the phenomenal character of occurrent non-sensory mental states like thoughts or wishes, and on the distinct subjective quality that goes along with thinking (see Bayne and Montague, 2011, for a good overview). Some philosophers claim that there is a *proprietary*, *distinctive*, and *individuating* phenomenology of higher cognitive processing that cannot be derived from sensory phenomenology, others deny this claim. For present purposes, I leave this controversial issue to the side (but see the point about predictive horizons in Section 3).

Let us introduce a working concept of ‘second-order mental action’. The satisfaction conditions of second-order mental actions are constituted by successfully influencing other mental actions or mental behaviours, first-order mental processes are the targets of second-order mental action. Examples of second-order mental action are the termination of an ongoing violent fantasy, but also the deliberate strengthening and sustaining of a spontaneously arising pleasant day-dream, the effortful attempt to make an ongoing process of visual perception more precise by selectively controlling the focus of attention, or — as in mental calculation and logical thought — the process of imposing a very specific abstract *structure* on a temporal sequence of inner events, of ‘conducting’ a symbolic train of thought (McVay and Kane, 2009). Philosophically, it is interesting to note how second-order mental actions are essential tools for achieving higher degrees of mental autonomy and self-determination; and also how many of them can be described as processes of computational resource allocation in the brain — for example, in the case of attentional agency, as an active optimization of precision expectations (Friston, 2010; Hohwy, 2013; Clark, 2015). However, an important distinction is the difference between *possessing* an ability (for example, the ‘tool’ of second-order mental action) and having an explicit *knowledge* that oneself possesses this ability. What Schooler and colleagues have provisionally termed ‘meta-awareness’ (Schooler *et al.*, 2011) is a necessary precondition for second-order mental action.

We may treat the preceding discussion as a first set of empirical, phenomenological, and conceptual constraints that any good philosophical theory should satisfy, and then ask: What exactly is autonomy at the mental level? First, because developmentally as well as phenomenologically AA clearly is the more basic form of epistemic mental agency, we need a subdoxastic account of autonomy here, one that does not presuppose rationality constraints, propositional attitudes, or access to some Sellarsian or other kind of ‘logical space of reasons’. Please recall how, above, I already pointed out that consciously experienced AA is theoretically important, because it is probably the earliest and simplest form of experiencing oneself as an autonomous epistemic agent. A second point of interest is that, at least in human beings, it not only causally enables high-level rational thought, but helps to constitute it: AA can exist without CA, but it is a necessary condition for CA. If we cannot control our attention, we cannot engage in rational, logically structured thought, but on the



other hand there are many self-conscious biological systems that, while not having the capacity for high-level rationality, can actively control and even become the object of their own attention (think of mirror-self-recognition in chimpanzees, bottlenose dolphins, or the Eurasian magpie). Second, we want a working concept of M-autonomy that facilitates interdisciplinary cooperation by being open to fine-grained functional analysis, yielding testable empirical predictions (e.g. the existence of a ‘self-representational blink’ following every single loss of autonomy, see Metzinger, 2013a).

‘Veto control’, or the capacity for intentional inhibition, may have to be the central semantic element in our new working concept of M-autonomy — simply because if you cannot terminate your very own activity, then you cannot be said to be autonomous in any interesting sense. This element can be empirically grounded, gradually refined, and may prove heuristically fruitful in guiding future research. Veto control is a manifestation of the capacity to voluntarily suspend or inhibit an action, and from a logical point of view it is a functional property which we do not ascribe to the brain, but to the person as a whole. Let us call the capacity in question ‘intentional inhibition’.<sup>5</sup> During a mind-wandering episode, we do not have this capacity, because we cannot actively suspend or inhibit our own mental activity. Recent empirical work reveals the dorsal fronto-median cortex (dFMC) as a candidate region for the physical realization of this very special form of purely mental second-order action.<sup>6</sup> It does not overlap with known networks for external inhibition, and its computational function may lie in predicting the social and more long-term individual consequences of a currently unfolding action, that is, in representing the action’s socially and temporally more distant implications for the organism.<sup>7</sup> There is a considerable amount of valuable neurobiological data on the physical substrates of intentional inhibition in human beings, and a number of them have already led to more abstract computational models of volitional control, action

---

<sup>5</sup> In adopting this terminological convention, I follow Marcel Brass (Brass and Haggard, (2007); an excellent and helpful recent review is Filevich, Kühn and Haggard (2012).

<sup>6</sup> See Kühn, Haggard and Brass (2009), Brass and Haggard (2007), Campbell-Meiklejohn *et al.* (2008). A helpful recent review of negative motor effects following direct cortical stimulation, listing the main sites of arrest responses and offering interesting discussion is Filevich, Kühn and Haggard (2012).

<sup>7</sup> This passage draws on Metzinger (2013a). See also Filevich, Kühn and Haggard (2012; 2013).

selection, and intention inhibition itself (Filevich, Kühn and Haggard, 2012; 2013; Campbell-Meiklejohn *et al.*, 2008; Kühn, Haggard and Brass, 2009; Brass and Haggard, 2007). These data are valuable not only for understanding the ‘back end’ of many mind-wandering episodes, but also for a more comprehensive theory of mental autonomy (for more, see Metzinger 2013a, Section 3.3).

Conceptually, many forms of mental self-control — like AA — presuppose exactly this ability for veto control, but are not directly guided by consciously represented reasons, explicit logical inferences, or arguments. Indeed, there is no need or even conceptual necessity to specify autonomy as *rational* self-control, because our capacity for rational self-control is only a special case of a more comprehensive, fundamental set of functional properties. First, rationality does not have to express itself in terms of explicit, symbolic reasoning processes using propositional data-formats (e.g. a Fodorian ‘language of thought’), but can be operationally defined as a property of some global input-output-function maximizing a specific fitness criterion. Second, there are more operational and empirically grounded models of autonomy, combining the notion of causal self-determination with independence from alternative causes, both inner and outer (see Seth, 2010, for the notion of ‘G-autonomy’ based on a formal analysis of Granger causality). For empirical research programmes on mind-wandering, such operational concepts are more likely to yield specific, testable hypotheses. Nevertheless, the notion of ‘rational mental self-control’ in the traditional sense remains important if we want to understand the *phenomenology* of high-level cognition and the normative components of our concept of ‘personhood’. Explicit rational self-control at the mental level cannot be reduced to veto control — on the contrary, the capacity for veto autonomy is only one of its centrally relevant constitutive conditions. Clearly, the capacity for inhibiting mental processes via second-order acts of vetoing without the involvement of quasi-conceptual or quasi-propositional representations is the more frequent and also more basic phenomenon, and hence also the more fundamentally relevant target for research. You can only be rational if you have the capacity for mental veto control, but you can achieve a high degree of mental autonomy without rational self-control.

This yields a working concept of M-autonomy as the ability to control the conscious contents of one’s mind in a goal-directed way, by means of attentional or cognitive agency. This ability can be a form of rational self-control, which is based on reasons, beliefs, and con-

ceptual thought, but it does not have to be. What is crucial is the ‘veto component’: being mentally autonomous means that all currently ongoing processes can in principle be suspended or terminated. Importantly, this does not mean that they actually *are* terminated, it just means that the ability, the functional potential, is given and that the person has knowledge of this fact. This point provides us with a third and equally important phenomenological constraint: if we only ‘tune out’, but do not ‘zone out’ — for example if we observe the spontaneous arising of memories or the beginning stages of a day-dream, or if we even voluntarily indulge in a fantasy while all the time knowing that we could terminate this inner activity at any instant — then we possess M-autonomy (see note #9). Call this the principle of ‘Autonomy by Phenomenal Self-Representation’: one can only deliberately and autonomously exert an ability if that ability is explicitly represented in one’s phenomenal self-model. Terminating a train of thought is one example of such an ability, detaching the focus of attention from a perceptual object is another. In sum, M-autonomy is the capacity for causal self-determination at the mental level. It is based on a complex and graded functional property, which comes in three major degrees: the phenomenally represented *knowledge* that *oneself* currently possesses this specific ability, executed *attentional* self-control, and *cognitive* self-control.

## **2. M-Autonomy, the First-Person Perspective, and Personhood**

For a human being, to possess a consciously experienced first-person perspective means to have acquired a very specific functional profile and a distinctive level of representational content in one’s currently active phenomenal self-model: it has, episodically, become a dynamic inner model of a *knowing self*. Representing facts under such a model creates a new epistemic modality. All knowledge is now accessed under a new internal mode of presentation, namely, as knowledge possessed by a self-conscious entity intentionally directed at the world. Therefore, it is *subjective* knowledge. This notion of a conscious model of oneself as an individual entity actively trying to establish epistemic relations to the world and to oneself, I think, comes very close to what we traditionally mean by notions like ‘subjectivity’ or ‘possession of a first-person perspective’. If we combine this observation with the concept of M-autonomy, then we can perhaps gain a fresh, empirically grounded, and conceptually enriched

perspective on traditional philosophical puzzles related to concepts like ‘perspectivalness’ and ‘personhood’.

Let us introduce a second conceptual instrument. The concept of an ‘epistemic agent model’, or EAM, refers to a specific type of conscious self-representation, a small subset of phenomenal self-models (PSMs).<sup>8</sup> This simply means that, at the level of conscious experience, the self is represented as something that either currently stands in an epistemic relation to the world, in the relation of knowing, thinking, actively guiding attention, or actively trying to understand what is going on in its environment; or, more abstractly, as an entity that has the *ability* to do so.<sup>9</sup> For any information processing system, to possess a first-person perspective means to operate under a specific kind of conscious self-representation, a PSM that portrays the system as an epistemic agent, as an entity that is actively searching for and optimizing its knowledge, for example by controlling its own high-level, quasi-symbolic processing as a cognitive agent (CA) or by actively sustaining and controlling the focus of attention (AA). This is what I call an EAM.<sup>10</sup> Again, having an EAM is a special case of

- 
- <sup>8</sup> A useful conceptual instrument to develop more fine-grained descriptions of the phenomenology of mind-wandering and the episodic reappearance of M-autonomy is the notion of a ‘phenomenal self-model’ (PSM; Metzinger, 2003a; 2006; 2008). A PSM is a conscious representation of the system as a whole, including not only global body representation (Metzinger, 2014; Blanke and Metzinger, 2009), but also psychological, social, and other potential personal-level properties. One central idea of the self-model theory (Metzinger, 2003a) is that, under standard conditions, a large part of the human PSM is ‘transparent’, because we are not able to experience it *as* a model and therefore fully identify with its representational content. Having an EAM is a special case of having a PSM.
- <sup>9</sup> This is not to say that we never purposefully engage in daydreams or that there are never situations in which we are mind-wandering while being passively aware of this fact. This is only to say that intentional episodes of daydreaming, to the extent that they do involve the phenomenology of AA and CA, thereby do not count as episodes of mind-wandering, which refer only to unintentional episodes of stimulus-independent thought. One advantage of the terminological solution proposed here is exactly that it enables a continuous description of real-world cases: as long as the EAM still represents the *ability* to become an active attentional or cognitive agent, we have M-autonomy. What has been termed ‘zoning out’ (unaware mind-wandering) and ‘tuning out’ (mind-wandering with awareness) in the empirical literature (Smallwood, McSpadden and Schooler, 2007, p. 524; 2008; Schooler *et al.*, 2011, p. 323) can be nicely captured by this conceptual distinction.
- <sup>10</sup> For details, see Metzinger (2003a, and 2006, Section 4). The philosophical notion of a ‘phenomenal model of the intentionality relation’ (PMIR) is directly related to the idea of dynamically integrating top-down control (e.g. by the fronto-parietal control network) with subpersonal, bottom-up components (e.g. a subset of activity in the default mode network) by creating an internal model of the whole organism as currently *being*

having a PSM, not all PSMs are EAMs. Empirically, it has been shown that human beings can enjoy a minimal form of self-consciousness without possessing an EAM (Blanke and Metzinger, 2009; Limanowski and Blankenburg, 2013). The transition from simple, bodily self-identification to the relevant, stronger form takes place when a system phenomenally represents itself as an entity capable of epistemic agency, or even as one currently exerting epistemic agency. If such a specific kind of self-model is in place, ongoing processes can be *embedded* into it, thereby creating the phenomenology of ownership (*my thought, my own* autobiographical memory, *my own* future planning). If these processes are additionally represented as control processes, as successful acts of exerting causal influence, they can now be consciously experienced as processes of *self-control* or instances of successful mental *self-determination*. An EAM is an instrument in what one might call ‘epistemic autoregulation’: it helps a self-conscious system in selecting and determining what it will know, and what it will not know. Yet, an epistemic agent model of this kind is not a little man in the head, but itself an entirely subpersonal process. During full-blown episodes of mind-wandering, we are not epistemic agents, neither as controllers of attentional focus nor as deliberate thinkers of thoughts, and we have forgotten about our agentic abilities. A first interim conclusion then is that what really takes place at the onset of a mind-wandering episode must be a collapse of the EAM.

It would perhaps be tempting to say that during such periods we have altogether lost the functional ability to control our own thought. I want to defend a more moderate, nuanced position: what we have lost is a specific form of *knowledge*, and not the ability itself, namely, conscious knowledge of our potential for second-order mental action. We are still persons, because we have the relevant potential. But we currently lack an explicit and globally available *representation* of an existing functional ability for active epistemic self-control — because we have not epistemically appropriated it. And that is exactly what an EAM does for us. But why do we then have the feeling that all of this cannot be an accurate phenomenological description of a very large portion of our conscious lives? Because we confuse our abstract,

---

*directed* at an object component, for example, by means of a well-ordered train of thought; see Smallwood *et al.* (2012). The PMIR would then be the conscious correlate of this process, the phenomenal experience of what was termed CA in the main text.

retrospective, and purely intellectual knowledge that, in principle, we had the critical mental ability all along with what actually was the case on the level of concrete, inner phenomenology: the absence of an EAM. As Franz Brentano (1874/1973, pp. 165f.) and much later Daniel Dennett (1991, p. 359) have pointed out, the representation of absence is not the same as the absence of representation.

It follows that in most cases the re-emergence of an EAM will have to be caused by an unconscious event, perhaps by chance, perhaps based on an *implicit* knowledge about the relevant potential, about an already existing ability. Whenever the dynamic process of creating and sustaining an EAM takes place, we also have a first-person perspective. AA is one specific example of having a consciously experienced first-person perspective. Its theoretical relevance consists in the fact that it is plausibly the simplest form of an EAM human beings can have. We still lack an empirically grounded theory of subjectivity, a model of the first-person perspective as a naturally evolved phenomenon (Metzinger, 2003a). But it is clear that having a first-person perspective is not a unitary but a graded phenomenon, and research on mind-wandering can make decisive contributions by functionally dissociating different levels. For example, we can see more clearly how attentional control is a necessary condition of personhood: you cannot engage in rational thought if you cannot control your own attention, because high-level epistemic autoregulation functionally presupposes low-level epistemic autoregulation.

Originally, the concept of a 'first-person perspective' is not much more than a visuo-grammatical metaphor. It has two different semantic components: the specific logic of the self-ascription of psychological properties using the first-person pronoun 'I', and the entirely contingent spatial geometry of our dominant sensory modality. Conscious vision of the human kind has a 'perspectival' geometrical structure, because it involves a single point of origin, namely, behind our eyes as phenomenally experienced. On a more abstract level, we may connect this phenomenological notion of an 'origin' constituting the centre of our internal model of reality with the origin of multimodal perceptual space ('here'), with self-location in a temporal order ('now'), and with the sensorimotor origins of *action* space, i.e. with the physical body ('embodiment'). Arguably, however, all of this only leads to a more or less minimal sense of selfhood (see Blanke and Metzinger, 2009), in which the subjectivity and perspectivalness of experience are mostly captured in an implicit or spatial sense. I think the concept of an EAM is interesting for any non-

trivial notion of subjectivity, because it isolates the origin of our *inner* space of action.

What about personhood? Clearly, an animal or artificial cognitive system could have a first-person perspective in this sense without counting as a person. On the other hand the potential for M-autonomy and the functional ability to (at least sometimes) operate under a conscious EAM are excellent candidates for criteria of personhood, which have the advantage of empirical grounding and hardware-independence at the same time. For the purposes of this paper, let us say that ‘personhood’ is a concept of social ontology. Personhood is constituted not in brains, but in societies — via a process in which human beings acknowledge each other as rational individuals possessing the capacity for moral thought and action. This makes the two concepts of M-autonomy and an EAM even more interesting: they potentially allow us to describe not only necessary conditions of personhood, but also the *transition* to personhood in a more fine-grained way. Human beings only *become* persons exactly by having the potential to phenomenologically identify with the content of an EAM, a step which on the sociocultural level causally enables relevant practices like linguistically ascribing person-status to themselves and mutually acknowledging each other as subjects of experience, as epistemic agents, and as morally sensitive, rational individuals. This led to a major expansion of our culturally structured cognitive niche and enabled the evolution of new forms of intelligence via a mutual scaffolding between all those individuals immersed in it. But why do we subjectively experience some of our cognitive processes *as* personal-level properties? There is a long story to be told here (Metzinger, 2003a; 2006; 2007; 2008), but the short answer is this: because they have been embedded into an EAM, which is currently active in our brain; and because we live in a normative sociocultural context in which we are now able to folk-psychologically describe and reciprocally acknowledge each other as rational individuals — a fact which then in turn influences introspective experience itself, turning the self-model into a person-model. From a functional perspective, M-autonomy dramatically expands our inner and outer space of possible behaviours, and one may speculate that perhaps it was exactly the emergence of an EAM which causally triggered the transition from biological to cultural evolution in our ancestors.

### 3. Conscious Thought is a Subpersonal Process

Before I present a simple, quantitative argument for the main claim of this paper, please follow me in considering an introductory example. It may help to further clarify and illustrate what has been said above. Imagine you are participating in a Buddhist-style silent retreat, an intensive course in mindfulness meditation. During the first three days your teacher instructs you to very precisely observe your breath as it comes and goes, but without in any way interfering with the respiratory process itself. Your task is to, whenever you have noticed an incoming thought or any other sort of distraction, gently bring back your attention to the bodily sensations going along with the rise and fall of your chest or abdomen, and to the sensation of the breath at the nostrils and the internal flow of air. Whenever you notice another attentional lapse, you simply return to your breath. But later, as the retreat progresses, you are instructed to become non-judgmentally aware of those incoming thoughts themselves, as they come and go, not identifying with or reacting to them. Now your task is to simply be present with whatever arises in your conscious mind.

We have two different tasks, and, at least initially, two different kinds of mental action, leading to two different inner situations. Given these two situations — what exactly is it that you are phenomenally representing? Let us ask: what is your conscious experience an experience *of*? I claim that in both cases you are representing physical processes in the body, you are experiencing not actions, but events, namely, chains of *subpersonal* events. The properties instantiated during these processes are not properties of the person as a whole. Let us first look at the intentional object of your introspective experience from a metaphysical perspective.

From the perspective of metaphysics, to gain meta-awareness of ongoing mind-wandering really is almost exactly like gaining meta-awareness of your breath. The introspective experience of breathing, as well as seemingly task-unrelated, phenomenologically spontaneous thoughts, are not personal-level psychological processes that are mysteriously correlated with or caused by some physical chain of events. The most parsimonious metaphysical interpretation of the relevant scientific data is that they are *identical* with functionally complex, but sub-global physiological processes in the biological body. In the case of mind-wandering, this physiological process is a specific, widely distributed pattern of neural activity, and it is now empirically plausible to assume that large parts of this pattern overlap



with activity in the default mode network (DMN; Buckner, Andrews-Hanna and Schacter, 2008; Christoff, 2012, Christoff *et al.*, 2009; Weissman *et al.*, 2006; Stawarczyk *et al.*, 2011; Andrews-Hanna *et al.*, 2010; Mantini and Vanduffel, 2012; Buckner and Carroll, 2007; Mason *et al.*, 2007; Spreng, Mar and Kim, 2009), but that it also extends to other functional structures like the rostrolateral prefrontal cortex, dorsal anterior cingulate cortex, insula, temporopolar cortex, secondary somatosensory cortex, and lingual gyrus (for a recent meta-analysis, see Fox *et al.*, 2015). What we introspectively represent are specific, as yet unknown abstract properties of the physical dynamics characterizing this pattern.

‘Sub-global’ or ‘local’, however, does not automatically imply ‘sub-personal’. A sub-global physiological process in the brain can *become* a personal-level process by being functionally integrated and represented within an EAM. Conversely, simply being identical with a ‘global’ process in the brain does not automatically imply being attributable to the person as a whole. What is required for the relevant shift from the subpersonal to the personal level is an epistemic appropriation at a specific level of phenomenal self-consciousness, the functional integration into an EAM (as explained in Section 2). The wandering mind does not meet this criterion, it is therefore sub-global *and* subpersonal. As the brain is a part of our body, any rational research heuristics targeting the neural correlates for the introspective phenomenology of breathing, or alternatively the critical subset of neural activity underlying mind-wandering, will therefore treat them as subpersonal, bodily processes. They have a long evolutionary history (Corballis, 2013; Lu *et al.*, 2012; Mantini and Vanduffel, 2012), and both of them clearly are constituted by dynamic, self-organizing chains of neural events that continuously and automatically unfold over time. They are not agentic processes implying explicit goal-selection, rationality constraints, etc. The postulation of a local, domain-specific identity is a tenable, coherent metaphysical interpretation of this fact. Whatever will figure as the *explanans* in a future scientific theory of mind-wandering or the phenomenology of breathing will therefore not be global properties of ‘the mind’ or the person as a whole, but specific microfunctional properties realized by the local physical dynamics underlying each episode of consciously experienced subpersonal cognitive processing. Therefore, if one adds the straightforward metaphysical assumption of a domain-specific identity (Bickle, 2013; McCauley and Bechtel, 2001) holding between the phenomenal states constituting episodes of mind-wandering and

what we are currently beginning to discover and incrementally isolate as their local, minimally sufficient neural ‘correlates’ (NCCs; e.g. Chalmers, 2000), then it seems obvious that mind-wandering simply is the phenomenal awareness of a local bodily process. What the Buddhist meditator attends to is activity in the NCC for mind-wandering, the dynamics of a local physical process.

But what then explains the marked phenomenological difference between those two different inner situations? On a more abstract, representationalist level of description we would say that attention has been directed to two different content levels in the conscious self-model, to certain aspects of the *body-model* and to the internal dynamics of the *cognitive self-model*. The crucial difference between the phenomenological profile of mindfully observing the breath and that of ‘being present with whatever arises in the mind’ can now be explained by the fact that only in the first case we find the functional property of information being made globally available through an interoceptive *receptor system*. Therefore, what Buddhists call ‘Anapanasati’ (or mindfulness of breathing) generates a *sensory* phenomenology of bodily self-representation. By contrast, as the human brain is devoid of any self-directed sensory channels or receptor systems, the relevant subset of neural activity in the NCC for mind-wandering cannot be informationally accessed through any perception-like causal links — although it, too, is a bodily process. Consequently, the phenomenology of cognition must necessarily be a *non-sensory* phenomenology — although it can of course be *about* possible sensory perceptions, fantasy worlds, linked to motor simulations, affectively toned, etc.

Put differently, what the cognitive self-model continuously *predicts* (Friston, 2010; Hohwy, 2013; Clark, 2015) are just much more abstract aspects of reality, in a wider temporal frame of reference, and not ongoing events on the sensory sheet. The PSM can be seen as an integrated global hypothesis about the state of the system in which it appears, constituted by a large number of individual predictions or sub-hypotheses, which are hierarchically structured and optimized at different timescales. A conscious self-model is therefore composed of different layers of expectations, in a continuous attempt of minimizing uncertainty and prediction error related to the system itself. Some layers continuously target causal regularities in shorter time-windows, some extract regularities relative to larger time-windows. In the words of Jakob Hohwy:

...the difference between percepts and concepts comes out in terms of a gradual movement from variance to invariance, via spatiotemporal scales of causal regularities. There is thus no categorical difference between them; percepts are maintained in detail-rich internal models with a short prediction horizon and concepts in more detail-poor models with longer prediction horizons. (Hohwy, 2013, p. 72)

What, then, determines the two different introspective phenomenologies of breathing and thinking? First, there will necessarily be different internal data-formats corresponding to either direct perceptual coupling sustained by receptor-driven input or to its absence, as we are clearly dealing with very different hierarchical levels in the self-model. Second, the different prediction horizons functionally characterizing these levels will lead to an embodied, fully situated, and perceptually coupled sense of presence in the first case, and to an 'unextended', much more disembodied (and potentially 'absent-minded') phenomenology in which the temporal succession of inner events is more salient, while at the same time spatial qualities as well as their deep sensorimotor origins have become almost unnoticeable. With regard to these more abstract content-layers of the human self-model, the facts that philosophers have frequently overlooked are, first, that non-agentive cognitive phenomenology is much more widespread than intuitively assumed, and second, that, conceptually, it often is not a personal-level process at all.

Before presenting some empirical evidence, let us remain with the illustrative example of mindfulness meditation to see a second, equally relevant, point more clearly. It is not about the metaphysics, but about the epistemology of conscious self-knowledge. One advantage of the concept of 'M-autonomy' is that it also offers a new understanding of what classical mindfulness meditation is: it is a systematic and formal mental practice of cultivating M-autonomy. Because mindfulness and mind-wandering are opposing constructs (Mrazek, Smallwood and Schooler, 2012), the process of losing and regaining meta-awareness can be most closely studied in different stages of classical mindfulness meditation (Hölzel *et al.*, 2011; Slagter, Davidson and Lutz, 2011). In the early stages of object-orientated meditation, there will typically be cyclically recurring losses of M-autonomy (see Hasenkamp *et al.*, 2012, fig. 1; Metzinger, 2013a), plus an equally recurring second-order mental action, namely the decision to gently but firmly bring the focus of attention back to the formal object of meditation, for example to interoceptive sensations associated with the respiratory process. Here, the

phenomenology will often be one of mental agency, goal-directedness, and a mild sense of effort. In advanced stages of so-called ‘open monitoring’ meditation, however, the aperture of attention has gradually widened, typically resulting in an effortless and choiceless awareness of the present moment as a whole. Whereas in beginning stages of object-orientated mindfulness practice the meditator identifies with an internal model of a mental agent directed at a certain goal-state (‘the meditative self’), meta-awareness of the second kind is typically described as having an effortless and non-agentive quality. In the first case an EAM is present, leading to a process that would still count as personal-level, whereas in the second case we have meta-awareness without an EAM. It is important to understand that these are distinct phenomenological state-classes. Interestingly, even the neural correlates pertaining to this difference between ‘trying to meditate’ and ‘meditation effortlessly taking place’ are already beginning to emerge (Garrison *et al.*, 2013).

From an epistemological point of view it is now interesting to note how the conceptual distinction between AA and CA either as functional or as phenomenal properties allows for the possibility of *hallucinating* epistemic agency. We might experience ourselves as autonomous mental subjects, but in some cases this might be an adaptive form of self-deception or confabulation (Hippel and Trivers, 2011; Pliushch and Metzinger, 2015). For example, if a subject during an experimental design involving mindfulness-based stress reduction regains meta-awareness (Hölzel *et al.*, 2011; Mrazek, Smallwood and Schooler, 2012) and describes the experience as ‘I have just realized that I was daydreaming and redirected my attention to the current moment and the physical sensations caused by the process of breathing!’, it may be false to assume that, functionally, the ‘realization’ was actually a form of AA or CA (see Schooler *et al.*, 2011, and Metzinger, 2013a, Section 3.3). What is subjectively described or experienced as a form of second-order mental action may sometimes not be a personal-level event at all, but a shift in the subpersonal self-model that is then misdescribed on the level of self-report, an auto-phenomenological *post hoc*-confabulation.<sup>11</sup> To consciously represent

---

<sup>11</sup> Let me point to a structural commonality with well-known problems in dream research, which may shed further light on the issue of what exactly it means that a mind-wandering episode *ends*. First, there is the phenomenon of ‘false awakening’, that is of realistic dreams of waking up (Windt, 2015; Windt and Metzinger, 2007; Green, 1994); second, current research interestingly shows that there are different levels of stages of

oneself as just having exerted a certain mental ability does not mean that one actually *had* this ability; the phenomenology of M-autonomy does not justify the claim that the functional property of M-autonomy was actually present. Claiming so would be a category mistake in which epistemic properties are ascribed to something that does not intrinsically possess them (Metzinger and Windt, 2014, p. 287; 2015, p. 7).

Regaining M-autonomy — a functional transition that in healthy people probably takes place many hundred times every day — seems to be a form of mental self-constitution, because a new type of conscious self-model is created, an EAM, which may later change global properties of the system as a whole (e.g. turning it into a subject of experience, or being recognized as a rational individual by other cognitive systems). You can certainly own the thoughts generated by a wandering mind without an EAM (phenomenologically they are still yours) even if the knowledge that you have the causal capacity for self-control is not consciously available, not represented on the level of your PSM. But representing yourself as a cognitive agent leads to the instantiation of a new phenomenal property. Let us call it ‘epistemic self-causation’: according to subjective experience — at the very moment of ‘coming to’ as it were — you actively constitute yourself as a thinker of thoughts. You are now consciously representing yourself *as currently representing*, as an individual entity creating new states of itself that are not just ‘real’ or bodily states, but states that might be true or false. You have intentional properties. As I have said in earlier work, having a first-person perspective means to dynamically co-represent the intentional relation itself *while* you represent, to operate under a model of reality containing the ‘arrow of intentionality’, which includes a conscious model of the self as directed at the world. The conscious experience of epistemic self-causation would then be a result of exactly such a continuous process of dynamical self-organization, a non-agentive process leading to a new functional level in the PSM. Importantly, this also suggests that rationalizing the immediately preceding, earlier episode as having *been* under one’s control may be a functionally necessary way of re-

---

becoming lucid in a dream (Noreika *et al.*, 2010; Voss *et al.*, 2013; Metzinger, 2013b). If there is an additional awareness of meta-awareness as just having been regained (i.e. a third-order meta-representation or second-order EAM), then the point made in the previous paragraph also applies: as such, this is just phenomenal experience, and not necessarily knowledge — we might always be introspectively self-deceived.

establishing and preserving internal coherence of the new conscious self-model, even if this process involves a retrospective confabulation.

Leading empirical researchers come to the same conclusion. Schooler and colleagues, referring to work by the late Daniel Wegner, point out that regaining meta-awareness may be accompanied by an illusion of control (Schooler *et al.*, 2011, Box 1; Wegner, 2002). Whenever we have this case, it seems that a specific new self-model has appeared: an autobiographical self-representation falsely depicting the last mental event as something that was self-controlled, an instance of deliberate causal self-determination at the mental level. This form of control is often described as an *auto-epistemic* form of self-control, as an instance of actively acquired self-knowledge or a sudden insight. Thus, a typical auto-phenomenological report may claim ‘I have just regained meta-awareness, because I just introspectively realized that I was lost in mind-wandering!’. Do we have reason to believe such claims? Is the reappearance of meta-awareness a subpersonal event or is it something in which global control and the conscious EAM actually played a decisive causal role?

Here, my positive proposal would be that we may actually be confronted with a functionally adequate form of self-deception, at least in many such cases: the re-emergence of an EAM, really triggered by unconscious events, may necessarily involve a confabulatory element (‘I generated this insight myself!’) in order to ensure the coherence of the *autobiographical* self-model over time. In order to be able to conceive of myself as an autonomous mental agent again, I simply must have been a mental agent in the preceding conscious moment too, because I had the ability, the potential, all the time. The transition must have been self-caused, because a ‘representational bridge’ has to be built to earlier instances of M-autonomy, thereby preserving the (virtual) transtemporal identity of the conscious, thinking self. I cannot *consciously* simulate myself as having *unconsciously* known about my ability for epistemic agency in the past. The onset of every fresh period of M-autonomous cognition may therefore, necessarily, involve an element of misrepresentation: if I want to consciously represent myself as *just now* having acquired the capability of causal self-determination, I need to integrate the (subpersonal) event of transition into the currently active PSM, endowing it with the phenomenal property of ownership and connecting it with earlier such events. If this is true, an illusory phenomenology of self-causation will be a necessary neurocomputational fiction in the construction of any

new EAM, terminating the mind-wandering episode which preceded it.

The onset of a mind-wandering episode, on the other hand, can be understood as a loss of M-autonomy, because it involves an unnoticed loss of mental self-control and epistemic agency, either on the level of attention or of cognition. As an unintentional form of mental behaviour it is not rationally guided, and while it is unfolding it cannot be terminated at will. Mind-wandering is a failure of causal self-determination at the level of mental content, and although it clearly has aspects that can be described as functionally adaptive, its overall performance costs and its negative effects on general, subjective well-being are obvious and have been well documented (for example, in terms of reading comprehension, memory, sustained attention, or working memory, *cf.* Mooneyham and Schooler, 2013, Table 1). It is an important and philosophically relevant contribution of research on mind-wandering to have demonstrated the ubiquity of the phenomenon and its effects (Smallwood and Schooler, 2015).

Let us therefore look at some empirical constraints, which any convincing philosophical theory of what, today, we still call ‘conscious thought’ must satisfy. We know that conscious mind-wandering is a process that can get completely out of control (Schupak and Rosenthal, 2009; Bigelsen and Schupak, 2011), but that can also come completely to rest, either in practitioners of mindfulness meditation (Mrazek, Smallwood and Schooler, 2012; Slagter, Davidson and Lutz, 2011) or following lesions to the medial frontal cortex (Damasio and van Hoesen, 1983). Under normal conditions, we spend 30–50% of our conscious waking lives mind-wandering (Kane *et al.*, 2007; Killingsworth and Gilbert, 2010; Schooler *et al.*, 2011). During these times we do not possess M-autonomy. If we assume a 16-hour day period, 40% of waking mind-wandering would amount to an average of 384 minutes, a period during which we are not autonomous mental subjects. NREM-sleep mentation and non-lucid dreaming clearly are also periods during which the functional property of M-autonomy is absent, although complex cognitive processes are taking place across all sleep stages (Windt, 2014; Wamsley, 2013; Fox *et al.*, 2013; Nielsen, 2000; Fosse, Stickgold and Hobson, 2001). They can be sampled and statistically evaluated, for example using a serial awakening paradigm (Noreika *et al.*, 2009; Siclari *et al.*, 2013). Although great progress has recently been made in isolating the neural correlates of dream lucidity (Dresler *et al.*, 2012; Voss *et al.*, 2009) and developing a more fine-grained conceptual taxonomy for different

kinds of lucidity (Noreika *et al.*, 2010; Voss *et al.*, 2013; Voss and Hobson, 2015), it remains clear that M-autonomy during the dream state is a very rare, and therefore negligible, phenomenon.

Adults spend approximately 1.5 to 2 hours per night in REM sleep (Hobson, 2002, pp. 77–79f.). NREM sleep yields similar reports during stage 1, other stages of NREM sleep are characterized by more purely cognitive/symbolic mentation. Clearly conscious thought during NREM sleep also lacks M-autonomy, because it is mostly confused, non-progressive, and perseverative. Whereas 81.9% of awakenings from REM sleep yield mentation reports, the incidence of reports following NREM awakenings lies at only 43% percent (Nielsen, 2000, p. 855). If we assume an average REM-time of 105 minutes, there will be an average of 86 minutes characterized by phenomenally represented, but subpersonal cognitive processing; 375 minutes of NREM sleep will yield roughly 161 minutes of conscious mentation, again, without M-autonomy. Assuming a waking period of 960 minutes, a very rough, first-order approximation is that human beings enjoy one sort of phenomenology or another for about 20 hours a day (1207 minutes; or about 84% of their daytime).

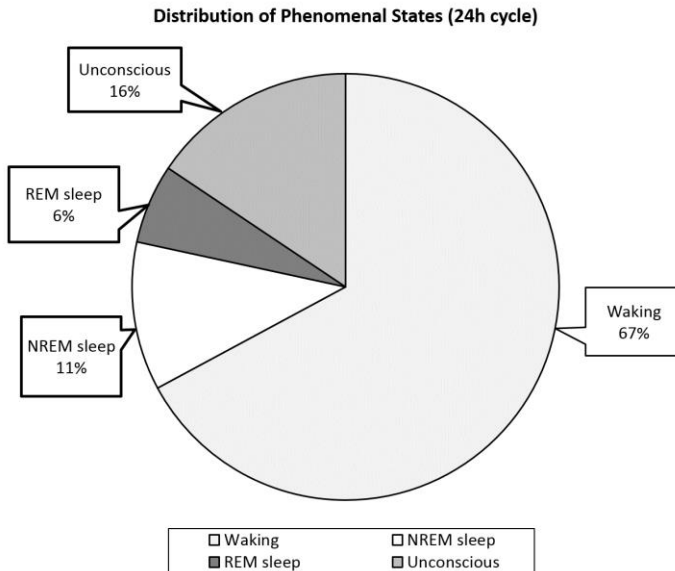


Figure 1. Distribution of conscious experience over the 24-hour-cycle.



However, healthy adults are only M-autonomous for 9.6 hours (576 minutes; or 40% of an average day). These are very conservative estimates. For example, they also exclude lifetime periods of illness, intoxication, or anaesthesia. In addition, there is evidence for extended periods in which human beings lose M-autonomy altogether. These episodes may often not be remembered and also frequently escape detection by external observers, as in ‘mind-blanking’ (Ward and Wegner, 2013). The same may also be true of periods of insomnia, in which people are plagued by intrusive thoughts, feelings of regret, shame, and guilt while suffering from dysfunctional forms of cognitive control, such as thought suppression, worry, depressive rumination, and counterfactual imagery (Schmidt, Harvey and van der Linden, 2011; Schmidt and van der Linden, 2009; Gay, Schmidt and van der Linden, 2011). We do not know when and how children actually acquire the necessary changes in their conscious self-model (Redshaw and Suddendorf, 2013), but we may certainly add the empirically plausible assumption that children only gradually acquire M-autonomy and that most of us likely lose it towards the ends of our lives.

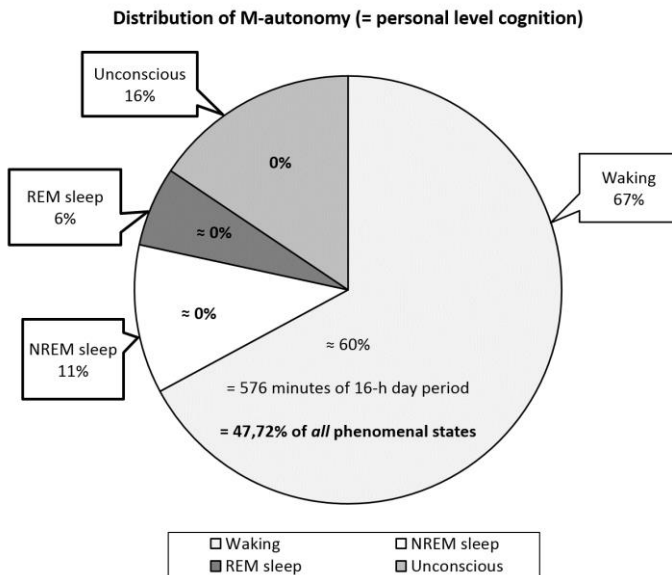


Figure 2. Distribution of M-autonomy over the 24-hour-cycle.

The first conclusion to be drawn from this first-order approximation is that, according to our preliminary working concept of M-autonomy, human beings, although phenomenally conscious, are not autonomous mental subjects for roughly two thirds of their conscious lifetime. A second, related conclusion is that conscious thought primarily and predominantly is an automatic subpersonal process, like respiration, heartbeat, or immune autoregulation — and that, at the conceptual level, we should do justice to this fact. It is empirically plausible to assume that a considerable part of our own cognitive phenomenology simply results from a frequent failure of executive control (McVay and Kane, 2009; 2010). I would claim that this actually is one of the most important functional and phenomenological characteristics of human self-consciousness, as a matter of fact, one of its most general, principal features: the almost constant presence of subpersonal and automatically generated mental activity (as generated by certain parts of the extended default mode network; Raichle *et al.*, 2001; Buckner, Andrews-Hanna and Schacter, 2008; Mantini and Vanduffel, 2012), in combination with a frequent inability of the executive-control system to shield primary-task performance off against interference from these subpersonal thought processes (Smallwood *et al.*, 2012). If I am right, autonomous cognitive self-control is an exception, not the rule.

To conclude, we may have to conceive the instantiation of a first-person perspective and certain necessary conditions of personhood as rare, graded, and dynamically variable properties of self-conscious cognitive systems — at least in our own case. I have proposed a background model of subjectivity as autonomous epistemic goal-selection at the mental level, with the EAM as the true origin of our consciously experienced first-person perspective. As the large majority of our mental activity is not driven by explicit, consciously available goal-representations and cannot, while it is unfolding, be inhibited, suspended, or terminated, we are not mentally autonomous subjects for about two thirds of our conscious lifetime. At the level of conscious mental activity, epistemic agency is the exception, not the rule. For human beings, epistemic agency can be differentiated into cognitive agency (CA; the ability to control goal-directed/task-related, deliberate thought) and attentional agency (AA; the ability to control the focus of attention). For most of their conscious lifetime, human beings are neither cognitive nor attentional agents, and they also lack an explicit phenomenal self-representation of themselves as currently *possessing* these abilities. Conceptually, most of our conscious activity must be characterized as a form of unintentional mental

behaviour. Therefore, two thirds of conscious thought can be described as a subpersonal process that functionally results from a continuously recurring loss of M-autonomy. However, I argued for a moderate interpretation of this fact, not in terms of a complete loss of the relevant ability, but only as an epistemic deficit, a lack of conscious self-knowledge: an absence of representation that is not represented as an absence.

From a philosophical perspective, mind-wandering is not a property of the person as a whole, but a local dynamics that is determined by a set of functional properties physically realized by a specific part of the brain. If it is in accordance with our theoretical interests, we may choose to describe this dynamics as a *representational* kind of dynamics. Then we can say that, internally, an individual representational token or event only *becomes* part of a personal-level process by being functionally integrated into and actively controlled with the help of a specific form of transparent conscious self-representation, the ‘epistemic agent model’ (EAM), and by being embedded into a highly specific sociocultural context. This context may be thought of as providing an external scaffolding for the stabilization of the EAM — for example, by enabling normative practices of mutually ascribing personhood to each other, or by realizing a linguistically structured cognitive niche in which the concept of a ‘person’ is continuously present as an instrument for social as well as mental self-representation. Here, one important conceptual distinction is the one between conscious self-representation of ongoing cognitive or attentional agency and a more implicit, passive representation of the *ability* to act as an epistemic agent, involving the more subtle phenomenology of knowing about the potential for mental action without actually realizing it. Being aware of this ability, also in others, suffices for the appearance of a first-person perspective.

In interdisciplinary discourse, it has now become a standard, and at times tiring, job for philosophers to tell neuroscientists that it is the *person* who thinks, and not the brain — a perennial job, it seems, because most neuroscientists never seem to really learn. The omnipresent and all too well-known mistake is the ascription of psychological predicates to parts of a person’s brain (e.g. ‘The prefrontal cortex plans actions’; ‘The premotor cortex decides on the initiation and organization of own movement sequences’; and so on). The conceptual error of ascribing a property that can only be ascribed to the whole entity to a part of it (called the ‘mereological fallacy’; see Bennett and Hacker, 2003, p. 72) often, but not necessarily,

accompanies the explanatory error of ascribing mental properties to subpersonal *explananda* (the ‘homunculus fallacy’; for a lucid discussion, see Drayson 2012, Section 2.2). Ironically, if what I have said above is correct, then at least some neuroscientists, in all their slightly indocile stubbornness, may actually have had a better intuition than philosophers: if for more than two thirds of our conscious lifetime ‘thought’ should better be described as an unintentional, sub-personal process, then most of the time it really is the *brain* that thinks — and not us.

### References

- Andrews-Hanna, J.R., Reidler, J.S., Huang, C. & Buckner, R.L. (2010) Evidence for the default network’s role in spontaneous cognition, *Journal of Neurophysiology*, **104**, pp. 322–335.
- Bayne, T. & Montague, M. (eds.) (2011) *Cognitive Phenomenology*, Oxford: Oxford University Press.
- Bennett, M.R. & Hacker, P.M.S. (2003) *Philosophical Foundations of Neuroscience*, Oxford: Blackwell.
- Bickle, J. (2013) Multiple realizability, in Zalta, E.N. (ed.) *The Stanford Encyclopedia of Philosophy, Spring 2013 Edition*, [Online], <http://plato.stanford.edu/archives/spr2013/entries/multiple-realizability/>.
- Bigelsen, J. & Schupak, C. (2011) Compulsive fantasy: Proposed evidence of an under-reported syndrome through a systematic study of 90 self-identified non-normative fantasizers, *Consciousness & Cognition*, **20**, pp. 1634–1648.
- Blanke, O. & Metzinger, T. (2009) Full-body illusions and minimal phenomenal selfhood, *Trends in Cognitive Sciences*, **13**, pp. 7–13.
- Brass, M. & Haggard, P. (2007) To do or not to do: The neural signature of self-control, *Journal of Neuroscience*, **27**, pp. 9141–9145.
- Brentano, F. (1874/1973) *Psychology from an Empirical Standpoint*, McAlister, L. (ed.), London: Routledge and Kegan Paul.
- Buckner, R.L. & Carroll, D.C. (2007) Self-projection and the brain, *Trends in Cognitive Sciences*, **11**, pp. 49–57.
- Buckner, R.L., Andrews-Hanna, J.R. & Schacter, D.L. (2008) The brain’s default network, *Annals of the New York Academy of Sciences*, **1124**, pp. 1–38.
- Campbell-Meiklejohn, D.K., Woolrich, M.W., Passingham, R.E. & Rogers, R.D. (2008) Knowing when to stop: The brain mechanisms of chasing losses, *Biological Psychiatry*, **63**, pp. 293–300.
- Carruthers, P. (2015) *The Centered Mind*, New York: Oxford University Press.
- Chalmers, D.J. (2000) What is a neural correlate of consciousness?, in Metzinger, T. (ed.) *Neural Correlates of Consciousness: Empirical and Conceptual Questions*, pp. 17–39, Cambridge, MA: MIT Press.
- Christoff, K. (2012) Undirected thought: Neural determinants and correlates, *Brain Research*, **1428**, pp. 51–59.
- Christoff, K., Gordon, A.M., Smallwood, J., Smith, R. & Schooler, J.W. (2009) Experience sampling during fMRI reveals default network and executive system contributions to mind wandering., *Proceedings of the National Academy of Sciences*, **106**, pp. 8719–8724.

- Clarke, A. (2015) *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford: Oxford University Press.
- Cohen, A.-L. (2013) Attentional decoupling while pursuing intentions: A form of mind wandering?, *Frontiers in Psychology*, **4**, p. 693.
- Corballis, M.C. (2013) Wandering tales: Evolutionary origins of mental time travel and language, *Frontiers in Psychology*, **4**, p. 485.
- Damasio, A.R. & van Hoesen, G.W. (1983) Emotional disturbances associated with focal lesions of the limbic frontal lobe, *Neuropsychology of Human Emotion*, **1**, pp. 85–110.
- Dennett, D.C. (1991) *Consciousness Explained*, Boston, MA: Little Brown.
- Dorsch, F. (2014) Focused daydreaming and mind-wandering, *Review of Philosophy and Psychology*, pp. 1–23, [Online], <http://link.springer.com/article/10.1007%2Fs13164-014-0221-4>.
- Drayson, Z. (2012) The uses and abuses of the personal/subpersonal distinction, *Philosophical Perspectives*, **26**, pp. 1–18.
- Dresler, M., Wehrle, R., Spoormaker, V.I., Koch, S.P., Holsboer, F., Steiger, A., *et al.* (2012) Neural correlates of dream lucidity obtained from contrasting lucid versus non-lucid REM sleep: A combined EEG/fMRI case study, *SLEEP*, **35**, pp. 1017–1020.
- Filevich, E., Kühn, S. & Haggard, P. (2012) Intentional inhibition in human action: The power of ‘no’, *Neuroscience and Biobehavioral Reviews*, **36**, pp. 1107–1118.
- Filevich, E., Kühn, S. & Haggard, P. (2013) There is no free won’t: Antecedent brain activity predicts decisions to inhibit, *PLoS ONE*, **8**, e53053.
- Fosse, R., Stickgold, R. & Hobson, J.A. (2001) Brain-mind states: Reciprocal variation in thoughts and hallucinations, *Psychological Science*, **12**, pp. 30–36.
- Fox, K.C.R., Nijeboer, S., Solomonova, E., Domhoff, G.W. & Christoff, K. (2013) Dreaming as mind wandering: Evidence from functional neuroimaging and first-person content reports, *Frontiers in Human Neuroscience*, **7**, p. 412.
- Fox, K.C.R., Spreng, R.N., Ellamil, M., Andrews-Hanna, J.R. & Christoff, K. (2015) The wandering brain: Meta-analysis of functional neuroimaging studies of mind-wandering and related spontaneous thought processes, *Neuroimage*, **111**, pp. 611–621.
- Friston, K. (2010) The free-energy principle: A unified brain theory?, *Nature Reviews Neuroscience*, **11**, pp. 127–138.
- Garrison, K.A., Santoyo, J.F., Davis, J.H., Thornhill, I.V., Thomas, A., Kerr, C.E. & Brewer, J.A. (2013) Effortless awareness: Using real time neurofeedback to investigate correlates of posterior cingulate cortex activity in meditators’ self-report, *Frontiers in Human Neuroscience*, **7** (440), pp. 1–9.
- Gay, P., Schmidt, R.E. & van der Linden, M. (2011) Impulsivity and intrusive thoughts: Related manifestations of self-control difficulties?, *Cognitive Therapy and Research*, **35**, pp. 293–303.
- Green, C.E. (1994) *Lucid Dreaming: The Paradox of Consciousness During Sleep*, Hove: Psychology Press.
- Hasenkamp, W., Wilson-Mendenhall, C.D., Duncan, E. & Barsalou, L.W. (2012) Mind wandering and attention during focused meditation: A fine-grained temporal analysis of fluctuating cognitive states, *Neuroimage*, **59** (1), pp. 750–760.
- Hippel, W. von & Trivers, R. (2011) The evolution and psychology of self-deception, *Behavioral and Brain Sciences*, **34** (1), pp. 1–56.

- Hobson, J.A. (2002) *Dreaming: An Introduction to the Science of Sleep*, Oxford: Oxford University Press.
- Hohwy, J. (2013) *The Predictive Mind*, Oxford: Oxford University Press.
- Hölzel, B.K., Lazar, S.W., Gard, T., Schuman-Olivier, Z., Vago, D.R. & Ott, U. (2011) How does mindfulness meditation work? Proposing mechanisms of action from a conceptual and neural perspective, *Perspectives in Psychological Science*, **6**, pp. 537–559.
- Horowitz, S.G., Braun, A.R., Carr, W.S., Picchioni, D., Balkin, T.J., Fukunaga, M., et al. (2009) Decoupling of the brain's default mode network during deep sleep, *Proceedings of the National Academy of Sciences*, **106**, pp. 11376–11381.
- Irving, Z.C. (2015) Mind-wandering is unguided attention: Accounting for the 'purposeful' wanderer, *Philosophical Studies*, pp. 1–25, [Online], <http://link.springer.com/article/10.1007%2Fs11098-015-0506-1>.
- Kane, M.J., Brown, L.H., McVay, J.C., Silvia, P.J., Myin-Germeys, I. & Kwapil, T.R. (2007) For whom the mind wanders, and when: An experience-sampling study of working memory and executive control in daily life, *Psychological Science*, **18**, pp. 614–621.
- Killingsworth, M.A. & Gilbert, D.T. (2010) A wandering mind is an unhappy mind, *Science*, **330**, p. 932.
- Klinger, E. (2013) Goal commitments and the content of thoughts and dreams: Basic principles, *Frontiers in Psychology*, **4**, p. 415.
- Kühn, S., Haggard, P. & Brass, M. (2009) Intentional inhibition: How the 'veto-area' exerts control, *Human Brain Mapping*, **30**, pp. 2834–2843.
- Limanowski, J. & Blankenburg, F. (2013) Minimal self-models and the free energy principle, *Frontiers in Human Neuroscience*, **7**, p. 547.
- Lu, H., Zou, Q., Gu, H., Raichle, M.E., Stein, E.A. & Yang, Y. (2012) Rat brains also have a default mode network, *Proceedings of the National Academy of Sciences*, **109**, pp. 3979–3984.
- Mantini, D. & Vanduffel, W. (2012) Emerging roles of the brain's default network, *Neuroscientist*, **19**, pp. 76–87.
- Mason, M.F., Norton, M.I., Van Horn, J.D., Wegner, D.M., Grafton, S.T., et al. (2007) Wandering minds: The default network and stimulus-independent thought, *Science*, **315**, pp. 393–395.
- McCauley, R.N. & Bechtel, W. (2001) Explanatory pluralism and heuristic identity theory, *Theory & Psychology*, **11**, pp. 736–760.
- McVay, J.C. & Kane, M.J. (2009) Conducting the train of thought: Working memory capacity, goal neglect, and mind wandering in an executive-control task, *Journal of Experimental Psychology: Learning, Memory & Cognition*, **35**, p. 196.
- McVay, J.C. & Kane, M.J. (2010) Does mind wandering reflect executive function or executive failure? Comment on Smallwood and Schooler (2006) and Watkins (2008), *Psychological Bulletin*, **136**, pp. 188–197.
- McVay, J.C. & Kane, M.J. (2013) Dispatching the wandering mind? Toward a laboratory method for cuing 'spontaneous' off-task thought, *Frontiers in Psychology*, **4**, p. 570.
- Metzinger, T. (ed.) (1995) *Conscious Experience*, Exeter: Imprint Academic.
- Metzinger, T. (2003a) *Being No One: The Self-Model Theory of Subjectivity*, Cambridge, MA: MIT Press.
- Metzinger, T. (2003b) Phenomenal transparency and cognitive self-reference, *Phenomenology and the Cognitive Sciences*, **2**, pp. 353–393.

- Metzinger, T. (2006) Précis: Being no one, *Psyche (Stuttg)*, **11**, pp. 1–35.
- Metzinger, T. (2007) Self models, *Scholarpedia*, **2**, p. 4174.
- Metzinger, T. (2008) Empirical perspectives from the self-model theory of subjectivity: A brief summary with examples, *Progress in Brain Research*, **168**, pp. 215–278.
- Metzinger, T. (2013a) The myth of cognitive agency: Subpersonal thinking as a cyclically recurring loss of mental autonomy, *Frontiers in Psychology*, **4** (931), [Online], <http://journal.frontiersin.org/article/10.3389/fpsyg.2013.00931/abstract>
- Metzinger, T. (2013b) Why are dreams interesting for philosophers? The example of minimal phenomenal selfhood, plus an agenda for future research, *Frontiers in Psychology*, **4** (746), [Online], <http://journal.frontiersin.org/article/10.3389/fpsyg.2013.00746/abstract>.
- Metzinger, T. (2014) First-order embodiment, second-order embodiment, third-order embodiment, in Shapiro, L. (ed.) *The Routledge Handbook of Embodied Cognition*, London: Routledge.
- Metzinger, T. & Windt, J.M. (2014) Die phänomenale Signatur des Wissens: Experimentelle Philosophie des Geistes mit oder ohne Intuitionen?, in Grundmann, T., Horvath, J. & Kipper, J. (eds.) *Die Experimentelle Philosophie in der Diskussion*, Berlin: Suhrkamp.
- Metzinger, T. & Windt, J.M. (2015) What does it mean to have an open mind?, in Metzinger, T. & Windt, J.M. (eds.) *OpenMIND*, Frankfurt am Main: MIND Group, [Online], <http://open-mind.net/papers/general-introduction-what-does-it-mean-to-have-an-open-mind/paperPDF>.
- Mooneyham, B.W. & Schooler, J.W. (2013) The costs and benefits of mind-wandering: A review, *Canadian Journal of Experimental Psychology*, **67**, pp. 11–18.
- Mrazek, M.D., Smallwood, J. & Schooler, J.W. (2012) Mindfulness and mind-wandering: Finding convergence through opposing constructs, *Emotion*, **12**, p. 442.
- Nielsen, T.A. (2000) A review of mentation in REM and NREM sleep: ‘Covert’ REM sleep as a possible reconciliation of two opposing models, *Behavioral & Brain Sciences*, **23**, pp. 851–866.
- Noreika, V., Valli, K., Lahtela, H. & Revonsuo, A. (2009) Early-night serial awakenings as a new paradigm for studies on NREM dreaming, *International Journal of Psychophysiology*, **74**, pp. 14–18.
- Noreika, V., Windt, J.M., Lenggenhager, B. & Karim, A.A. (2010) New perspectives for the study of lucid dreaming: From brain stimulation to philosophical theories of self-consciousness, *International Journal of Dream Research*, **3**, pp. 36–45.
- Plushch, I. & Metzinger, T. (2015) Self-deception and the dolphin model of cognition, in Gennaro, R.J. (ed.) *Disturbed Consciousness: New Essays on Psychopathology and Theories of Consciousness*, Cambridge, MA: MIT Press.
- Raichle, M.E., MacLeod, A.M., Snyder, A.Z., Powers, W.J., Gusnard, D.A. & Shulman, G.L. (2001) Inaugural article: A default mode of brain function, *Proceedings of the National Academy of Sciences*, **98**, pp. 676–682.
- Redshaw, J. & Suddendorf, T. (2013) Foresight beyond the very next event: Four-year-olds can link past and deferred future episodes, *Frontiers in Psychology*, **4**, p. 404.

- Samann, P.G., Wehrle, R., Hoehn, D., Spoormaker, V.I., Peters, H., Tully, C., *et al.* (2011) Development of the brain's default mode network from wakefulness to slow wave sleep, *Cerebral Cortex*, **21**, pp. 2082–2093.
- Schmidt, R.E. & van der Linden, M. (2009) The aftermath of rash action: Sleep-interfering counterfactual thoughts and emotions, *Emotion*, **9**, pp. 549–553.
- Schmidt, R.E., Harvey, A.G. & van der Linden, M. (2011) Cognitive and affective control in insomnia, *Frontiers in Psychology*, **2**, p. 349.
- Schooler, J.W., Smallwood, J., Christoff, K., Handy, T.C., Reichle, E.D. & Sayette, M.A. (2011) Meta-awareness, perceptual decoupling and the wandering mind, *Trends in Cognitive Sciences*, **15**, pp. 319–326.
- Schupak, C. & Rosenthal, J. (2009) Excessive daydreaming: A case history and discussion of mind wandering and high fantasy proneness, *Consciousness & Cognition*, **18**, pp. 290–292.
- Seth, A.K. (2010) Measuring autonomy and emergence via granger causality, *Artificial Life*, **16**, pp. 179–196.
- Siclari, F., La Rocque, J.J., Postle, B.R. & Tononi, G. (2013) Assessing sleep consciousness within subjects using a serial awakening paradigm, *Frontiers in Psychology*, **4**, p. 542.
- Slagter, H.A., Davidson, R.J. & Lutz, A. (2011) Mental training as a tool in the neuroscientific study of brain and cognitive plasticity, *Frontiers in Human Neuroscience*, **5**, p. 17.
- Smallwood, J. (2013) Distinguishing how from why the mind wanders: A process-occurrence framework for self-generated mental activity, *Psychological Bulletin*, **139**, p. 519.
- Smallwood, J., McSpadden, M. & Schooler, J.W. (2007) The lights are on but no one's home: Meta-awareness and the decoupling of attention when the mind wanders, *Psychonomic Bulletin & Review*, **14**, pp. 527–533.
- Smallwood, J., McSpadden, M. & Schooler, J.W. (2008) When attention matters: The curious incident of the wandering mind, *Memory and Cognition*, **36** (6), pp. 1144–1150.
- Smallwood, J., Brown, K., Baird, B. & Schooler, J.W. (2012) Cooperation between the default mode network and the frontal-parietal network in the production of an internal train of thought, *Brain Research*, **1428**, pp. 60–70.
- Smallwood, J. & Schooler, J.W. (2015) The science of mind wandering: Empirically navigating the stream of consciousness, *Annual Review of Psychology*, **66**, pp. 487–518.
- Spreng, R.N., Mar, R.A. & Kim, A.S.N. (2009) The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: A quantitative meta-analysis, *Journal of Cognitive Neuroscience*, **21**, pp. 489–510.
- Stawarczyk, D., Majerus, S., Maquet, P., D'Argembeau, A. & Gilbert, S. (2011) Neural correlates of ongoing conscious experience: Both task-unrelatedness and stimulus-independence are related to default network activity, *PLoS ONE*, **6**, e16997.
- Vanhoudenhuyse, A., Noirhomme, Q., Tshibanda, L.J.-F., Bruno, M.-A., Boveroux, P., Schnakers, C., *et al.* (2010) Default network connectivity reflects the level of consciousness in non-communicative brain-damaged patients, *Brain*, **133**, pp. 161–171.



- Voss, U., Holzmann, R., Tuin, I. & Hobson, J. (2009) Lucid dreaming: A state of consciousness with features of both waking and non-lucid dreaming, *Sleep*, **32**, p. 1191.
- Voss, U., Schermelleh-Engel, K., Windt, J., Frenzel, C. & Hobson, A. (2013) Measuring consciousness in dreams: The lucidity and consciousness in dreams scale, *Consciousness & Cognition*, **22**, pp. 8–21.
- Voss, U. & Hobson, A. (2015) What is the state-of-the-art on lucid dreaming?, in Metzinger, T. & Windt, J.M. (eds.) *OpenMIND*, Frankfurt am Main: MIND Group, [Online], [http://open-mind.net/papers/what-is-the-state-of-the-art-on-lucid-dreaming-recent-advances-and-questions-for-future-research/at\\_download/paperPDF](http://open-mind.net/papers/what-is-the-state-of-the-art-on-lucid-dreaming-recent-advances-and-questions-for-future-research/at_download/paperPDF).
- Wamsley, E.J. (2013) Dreaming as an extension of waking conscious experience and a tractable problem for cognitive neuroscience, *Frontiers in Psychology*, **4**, p. 637.
- Ward, A.F. & Wegner, D.M. (2013) Mind-blanking: When the mind goes away, *Frontiers in Psychology*, **4**, p. 650.
- Wegner, D.M. (2002) *The Illusion of Conscious Will*, Cambridge, MA: MIT Press.
- Weissman, D.H., Roberts, K.C., Visscher, K.M. & Woldorff, M.G. (2006) The neural bases of momentary lapses in attention, *Nature Neuroscience*, **9**, pp. 971–978.
- Windt, J.M. (2014) *Dreaming: A Conceptual Framework for Philosophy of Mind and Empirical Research*, Cambridge, MA: MIT Press.
- Windt, J.M. (2015) *Dreaming: A Conceptual Framework for Philosophy of Mind and Empirical Research*, Cambridge, MA: MIT Press.
- Windt, J.M. & Metzinger, T. (2007) The philosophy of dreaming and self-consciousness: What happens to the experiential subject during the dream state?, in *Praeger Perspectives: The New Science of Dreaming*, Vol. 3., *Cultural and Theoretical Perspectives*, pp. 193–247, Westport, CT: Praeger Publishers/Greenwood Publishing Group.