

aeon



# Are you sleepwalking now?

Given how little control we have of our wandering minds, how can we cultivate real mental autonomy?

*Thomas Metzinger*

Imagine you are standing on the prow of a sailboat, watching a school of dolphins leaping left and right. When travelling long distances, jumping saves dolphins energy, because there's less friction in the air than in the water below. It also seems to be an efficient way to move rapidly and breathe at the same time. Typically, the animals will alternate long, ballistic jumps with bouts of swimming underwater, close to the top, for about twice the length of the leap – a spectacular, high-speed, surface-piercing

display sometimes known as ‘porpoising’.

These cetacean acrobatics are a fruitful metaphor for what happens when we think. What most of us still call ‘our conscious thoughts’ are really like dolphins in our mind, jumping briefly out of the ocean of our unconscious for a short period before they submerge themselves once again. This ‘dolphin model of cognition’ helps us to understand the limits of our awareness. For example, the windows of time in which these leaps into consciousness unfold (as well as subsequent ‘underwater’ processing) vary hugely. And similar to the way that dolphins break the surface of the water, thoughts often cross the border between conscious and unconscious processing, and in both directions. Sometimes individual dolphins are so close to the surface that they can be half in and half out of the water; you might actually be able to learn how to spot them right before they jump, just as you can learn to identify subtle, semi-conscious patterns before they manifest as full-blown thoughts and feelings. There might even be more than one dolphin: in all likelihood, there’s a whole race going on between our thoughts, a continuous inner competition for the focus of attention and for what finally seizes control over our behaviour.

The point is that the mental contents available to us via introspection are nothing more than momentary flashes of automatic cognitive processing, grinding away beneath the waves of our awareness most of the time. This raises a strange question: who is the ‘us’, standing at the prow and watching these dolphin-thoughts scoot past? Philosophers of mind often fall into the trap of assuming that goal-directed, rational thought is the paradigmatic case of conscious cognition. But if *we* are only ever partly aware of what is happening in our own minds, surely *we* can’t be in absolute command of our thoughts, let alone causing them? Is it *ever* possible to distinguish mental *actions*, which we can direct and select, from the more general category of mental events, which simply happen to us? In what sense are we ever genuinely [mental agents](https://philpapers.org/archive/METTPO-14.pdf) [, capable of acting freely](https://philpapers.org/rec/METTPO-14), as opposed to being buffeted by forces beyond our control?

One of the most exciting recent [research](https://www.blogs.uni-mainz.de/fb05philosophieengl/files/2017/12/Metzinger_Mind_Wandering_OHST_2017_penultimate_draft.pdf) in neuroscience and experimental psychology is [mind-wandering – the study](http://www.annualreviews.org/doi/abs/10.1146/annurev-psych-010814-015331) of spontaneous or task-unrelated thoughts. Its results have radical implications for politics, education and morality. If we consider the empirical findings, we arrive at a surprising result of profound philosophical significance: cognitive control is the exception, while its absence is the rule. Much of the time we like to describe some foundational ‘self’ as the initiator or cause of our actions, but this is a pervasive myth

<https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00931/full>. In fact, we only resemble something *like* this for about a third of our conscious lifetime. We don't exactly know when and how children first learn to do it, and it's plausible that many of us gradually lose it towards the end of our lives. As far as our inner life is concerned, the science of mind-wandering implies that we're only rarely autonomous persons.

Mind-wandering research suggests that we need to get rid of naive, black-and-white distinctions such as 'free will' versus 'determinism', 'conscious' versus 'unconscious', and what philosophers call 'personal' versus 'sub-personal' processes (roughly, accounts of cognition that look at the whole person's reasons and beliefs, versus those based on biological or physiological functions). As the dolphin story hints, human beings are not *Cartesian* <https://aeon.co/ideas/descartes-was-wrong-a-person-is-a-person-through-other-persons> egos capable of complete self-determination. Nor are we primitive, robotic automata. Instead, our conscious inner life seems to be about the management of spontaneously emerging mental behaviour. Most of what populates our awareness unfolds automatically, just like a heartbeat or autoimmune response, but it can still be guided to a greater or lesser degree.

The really interesting question then becomes: *how* do various thoughts and actions 'surface', and what's the mechanism by which we corral them and make them our own? We ought to probe how our organism turns different sub-personal events into thoughts or states that appear to belong to 'us' as a whole, and how we can learn to control them more effectively and efficiently. This capacity creates what I call *mental autonomy* <http://www.ingentaconnect.com/content/imp/jcs/2015/00000022/F0020011/art00013?crawler=true>, and I believe it is the neglected ethical responsibility of government and society to help citizens cultivate it.

**T**he mind wanders more frequently than most of us think – several hundred times a day and up to 50 per cent of our waking life, in fact. We pay a high price for this mental peregrination. Scientific studies have shown that the spontaneously wandering mind has a negative impact on text comprehension and success at school, as well as on learning, attention, working memory, mathematical ability, and our capacity to drive a car safely. Daydreaming can also make us *unhappy* <http://science.sciencemag.org/content/330/6006/932>. Someone who loses contact with the present, because she zones out repeatedly into the future or the past, generally feels worse than someone who can keep her attention focused on what is happening now. There are also various types of insomnia, intoxication, light anaesthesia or illness (such as fever dreams or depressive rumination) in which we are in a sort of helpless twilight state, plagued by constantly recurring thoughts we cannot stop.

Not all forms of mental absence are the same, however. There's preliminary evidence that spontaneous thoughts play an important role in recovering from bad experiences, as well as in autobiographical planning, creative problem-solving, goal-directed thinking, and perhaps even in deeper forms of self-reflection. In fact, mind-wandering can often be seen as a process of emotional self-regulation. I have put forward the idea that this involves a semi-automatic juggling or switching between predictive horizons about what will happen to us next. The wandering mind is like a monkey, swinging from branch to branch across an inner emotional landscape. It will flee from unpleasant perceptions and feelings, and try to reach a state that feels better. If the present moment is unattractive or boring, then of course it's more pleasant to be planning the next holiday or drifting away into a romantic fantasy.

Clearly, we need to carve up the concept of mind-wandering with a bit more precision. Let's look at what's happening neurologically during these episodes. A multitude of empirical studies show that areas of our brain responsible for the wandering mind overlap to a large extent with the so-called *default-mode network* (DMN). This is a large network in our brain that typically becomes active during periods of rest, when attention is directed to the inside. The DMN is engaged during daydreams, for instance, as well as unbidden memories, or when we are thinking about ourselves and the future. However, as soon as a concrete task must be done, this part of our brain is deactivated and we concentrate immediately on a solution to the current problem.

My view is that the mind-wandering network and the DMN basically serve to keep our sense of self stable and in good shape. Like an automatic maintenance program, they constantly generate new stories, weaving back and forth between different time-horizons, each micro-narrative contributing to the illusion that we are actually the same person over time. Like nocturnal [dreaming <https://aeon.co/essays/why-are-dreams-such-potent-vehicles-for-the-supernatural>](https://aeon.co/essays/why-are-dreams-such-potent-vehicles-for-the-supernatural), mind-wandering also appears to be a process by which our brain and body consolidate our long-term memory and stabilise specific parts of what I call the '[self-model <http://www.scholarpedia.org/article/Self\\_models>](http://www.scholarpedia.org/article/Self_models)'.

I should come clean at this point and confess that I don't believe in any such entity or thing as 'the self'. At best, we have an inner image or representation of ourselves as a whole, made up of many functional modules and layers. At its most basic, this self-model is [based <https://aeon.co/ideas/our-illusory-sense-of-agency-has-a-deeply-important-social-purpose>](https://aeon.co/ideas/our-illusory-sense-of-agency-has-a-deeply-important-social-purpose) on an internal model of the body, including affective and emotional states, and grounded in inner-body perceptions such as gut feelings, heartbeat, breath, hunger or thirst. On another, higher layer, the self-model reflects a person's relationships to other people, ethical and cultural norms, and sense of self-worth. But in order to create a robust connection between the social and biological levels, the self-model fosters the illusion of transtemporal identity – the belief that we

are a whole and persisting entity based on the narrative our brain tells itself about 'our' past, present and future. (I think that it was exactly the impression of transtemporal identity that turned into one of the central factors in the emergence of large human societies, which rely on the understanding that it is *I* who will be punished or rewarded in the future. Only as long as we believe in our own continuing identity does it make sense for us to treat our fellow human beings fairly, for the consequences of our actions will, in the end, always concern us.)

## The self-as-agent is just a useful fiction, a neurocomputational artefact of our evolved self-model

But don't lose sight of the fact that all this modelling is just a convenient trick our organism plays on itself to enhance its chances of survival. We must not forget that the *phenomenal* realm (how we subjectively experience ourselves) is only a small part of the *neurobiological* one (the reality of the creatures we actually are). There's no little person in our head, only a set of dynamical, self-organising processes at play behind the scenes. Yet it seems like these processes often function by creating self-fulfilling prophecies; in other words, we have an identity because we convince ourselves we have one. Humans have evolved to be a bit like method actors, who need to really imagine and believe they *are* a particular character in order to perform effectively on stage. But just as there is no 'real' character, there's also no such thing as 'a self', and probably nothing like an immortal soul either.

Rather, one of the main functions of the self-model is how it lets our biological organism [predict <https://aeon.co/essays/consciousness-is-not-a-thing-but-a-process-of-inference>](https://aeon.co/essays/consciousness-is-not-a-thing-but-a-process-of-inference), and thereby control, the sensory consequences of our actions. That produces what's called our *sense of agency*. When we move an arm to grasp a glass or throw a ball, we need to anticipate how these movements will feel ahead of time in order for them to be successful. To come up with this prediction, and to minimise ugly surprises that could kill or harm us, we must develop a good explanation of the underlying cause of whatever we're experiencing. But since the real cause – unconscious, sub-personal processes, such as synapses firing – can't be represented within the workspace of our consciousness, the brain tells itself something else  [<https://aeon.co/essays/psychedelics-work-by-violating-our-models-of-self-and-the-world>](https://aeon.co/essays/psychedelics-work-by-violating-our-models-of-self-and-the-world): there must be *a self* acting so as to make all these thoughts and actions occur! The conscious experience of volition and agency are simple and elegant inferences to the best explanation. So when I close my fingers around the stem of a wineglass or feel the rough surface of a tennis ball in my hand, I infer that I must be an agent who is capable of originating, controlling and owning all these events.

The sense of agency involves three major components: a subjective, inner experience that you are the initiator or cause of your thoughts and bodily actions (what is called ‘ultimate origination’); the sustained impression of control or carrying through an action over a certain period of time; and a sense of ownership, the robust feeling that those thoughts and bodily actions belong to you. By contrast, a person suffering from schizophrenia might lack any or all of these things, because she is not able to integrate representations of her thoughts or actions into her self-model properly. Some patients report feeling like alien ideas are being implanted in their mind, or like mechanical puppets or robots, as if their body movements were controlled by an alien force.

However, even if you *do* experience yourself *as* an agent, that doesn’t mean you fundamentally *are* one. In the physical world, there is no such thing as ultimate origination. Science is what explains why you think and behave the way you do, not some pre-existing agential self. But just like a method actor can’t focus on the fact that she’s acting, our biological organism is usually unable to experience our self-model *as* a model. Instead, we tend to identify <https://aeon.co/essays/the-hard-problem-of-consciousness-is-a-distraction-from-the-real-one> with its *content*, just as the actor identifies with the character. The more we achieve a high degree of predictability over our behaviour, the more tempting it is to say: this is *me*, and I *did* this. We tell ourselves a brilliant and parsimonious causal story, even if it’s false from the third-person perspective of science. Empirically speaking, the self-as-agent is just a useful fiction or hypothesis, a neurocomputational artefact of our evolved self-model.

On the level of the brain, this process is a truly amazing affair, and a major achievement of evolution. But if we look at the resulting conscious experience from the *outside*, and on the level of the whole person, the brain’s mini-narrative also appears as a misrepresentation, slightly complacent, a bit grandiose, and ultimately delusional. Agency on the level of thought is really a ‘surface’ phenomenon, produced by the fact that the underwater, unconscious causal precursors are simply unknown to us. Even if we sometimes reach what resembles the rationalist ideal, we probably do so only sporadically, and the notion of controlled, effortful thinking is probably a very bad model of conscious thought in general. Our conscious mental activity is usually an unbidden, unintentional form of behaviour. Yet somehow the tourist on the prow begins to experience herself as an omnipotent magician, making dolphins come into existence out of the blue, and jump at her command.

**T**he self might not be a Cartesian agent that causes thought or action, but perhaps there are other ways for the organism as a whole to shape what happens in its mental life. We can’t get off the ship, let alone summon dolphins from nowhere, but perhaps we can choose where to look.

Among other things, this would equate to: the ability to impose rules on one's own mental behaviour; actively control the focus of attention; explicitly select mental goals; guide the flow of thought in accordance with reason and logic; and, most importantly, to intentionally end an ongoing mental process. These purposeful, inner actions are a large subset of the broader domain of mental behaviour or events (although it's still unclear how exactly they can be made amenable to scientific explanation). The capacity for such actions are what I referred to as mental autonomy, and they allow us to attain variable degrees of inner self-control.

We're familiar with the idea of autonomy over our actions in the outer realm, such as when we control our bodily movements and impulses ('I *will* exercise tomorrow; I will *not* eat that cookie.') According to subjective experience, however, there are not only bodily actions, but also mental ones. These might include actively re-directing attention to your breath in meditation, deliberately paying attention to a person's face in front of you, trying to retrieve visual images from your memory, logical thinking, or engaging in mental calculation.

Note that deliberately *not* acting is as important here as *acting*. The defining feature of autonomy in both the inner and outer realms is *veto control*, the power to inhibit, suspend or terminate ongoing actions. Someone who suffers from logorrhoea and cannot stop his outer flow of words will soon be unable to communicate with other human beings at all (a condition not unknown among philosophers). Likewise, those who can't get in touch with their own inner silence ultimately lose contact with themselves and soon won't be able to think clearly. If you sit down with the goal of writing a difficult text, or finding an answer to a challenging question, you must be able to resist thinking about your lunch, mulling over the last conversation you had with your mother, or wondering what's on the news. Anyone who can withstand or postpone such temptations must be capable of focusing on what they want to achieve in the long run, as a whole person, at the expense of other niggling demands or desires that impinge on their awareness. But if you lose focus for just a moment, your thought process is immediately hijacked by an aggressive little inner demon screaming 'Attend to me!' or 'Think me!', and your mind begins to wander.

A specific layer of the self-model is of central importance here. I call it the 'epistemic agent model' – the bit that allows us to have the feeling 'I am a knowing self; I know that I know.' This is the true origin of our first-person perspective. It's created by predictions about what the organism *can* and *will* know in the future, and helps us to continuously improve our model of reality. Think of what it's like when you start a new job: you quickly identify which new skills and pieces of knowledge you need to acquire, and set about filling the blanks. This is the epistemic-agent model in action. It's what lets us select the targets for our knowledge, permits us to actively focus our attention on certain problems and ideas, and decide what we want to understand

better. Crucially, the epistemic-agent model entails a specific kind of self-knowledge that sits at the heart of mental autonomy: for successfully exercising an ability entails not only being able to do it, but also *knowing* you have that capacity.

## What if your (triumphant) thought ‘I have just noticed that my mind has strayed!’ is itself an illusion of control?

Now we can see mind-wandering for what it really is: a transient loss of mental autonomy, via the loss of the epistemic-agent model. A daydream just *happens* to you – there is ownership, but no control over the event. It is not something you *do*, but something in which you ‘lose yourself’. You have forgotten a specific kind of self-knowledge, the ability to terminate a train of thought and to choose what it is you want to know. You might daydream *about* being a knowing self, but right now you have lost all awareness of your own power to put an end to the process.

Having a high degree of mental autonomy doesn’t mean you must stop what you’re doing right then and there. If you are deeply relaxed among nature, or sitting quietly in meditation, for example, then you might be able to rest peacefully in the present moment with a clear and silent mind, while still implicitly knowing that you could always suspend an incoming thought or get up and walk away at any point. As long as you have the capacity for self-control, and are aware of this fact, then your epistemic-agent model remains intact, without you having to actually exert any effort.

Just as you can experience autonomy in meditation, the same is (sometimes) true while you sleep. Dreaming can be seen as a kind of conscious thought but, unlike waking states, we usually possess no control over our thoughts or attention. Dreams appear to be a sort of unconstrained way of modelling the world. Sometimes, though, it’s possible to regain mental autonomy when we dream, in so-called ‘lucid dreaming’. When becoming lucid in a dream, a stable ‘knowing self’ appears. I think the epistemic-agent model is what gets added in the transition to lucidity and so overlaps with the part that gets shut down when we lose mental autonomy during the day. But it’s a genuine question as to whether human beings are actually capable of ‘lucid waking’ at all. How do you really know that you are *awake* right now?

Here’s one of the most personally unsettling ideas I discovered in the scientific literature on mind-wandering. In so-called ‘illusions of control’, human beings can hallucinate a role in chance events in which they’ve played no causal role. For example, people feel that if they choose tickets in a lottery game themselves, then they have a higher chance of winning than if someone else picks for them, or they actually throw a dice harder to get a higher number. People who ‘predicted’ coin tosses correctly in the beginning of a series begin to think that their performance would



improve with practice, and that distraction would impair their performance.

Now, imagine you're meditating when you notice that you've had an attentional lapse. Next, you regain meta-awareness of what your mind is doing, and return to your breath. What if your short (perhaps even slightly triumphant) thought 'Ah, I have just noticed that my mind has strayed!' is itself an illusion of control? Probably it wasn't *you* who noticed. Rather, the meditator's self-model might have quickly re-created the unexpected inner experience of 'coming to' as an achievement of the self, a sudden insight that was self-caused. But there never was a self that noticed anything.

Perhaps even mental autonomy, then, is just a very sophisticated trick that the self-model is playing on us, and doesn't involve a genuine form of predictive control. Perhaps mental autonomy plays *no* real causal role in our thought processes. If so, the very idea of autonomous mental action would seem to be a mere romantic fantasy, and we'd need to eliminate it entirely. Even high-level cognitive control might be just another form of mental sleepwalking. Let's call this option 'Radical Neuro-Buddhism', and come back to it at the end.

Such empirically grounded insights as these are examples of what I call the 'naturalistic turn in the image of humankind'. Genetics, cognitive neuroscience, evolutionary psychology and contemporary philosophy of mind are offering us a new image of ourselves based on an increasingly detailed understanding of our deepest cognitive structures. Whether we like it or not, we need to start looking at our mental abilities as natural properties with a biological history of their own. These features can be explained using the methods of science and, in principle, controlled technologically – maybe even implemented on non-biological carrier systems. In this wider context, it's easy to see how superrational future AI systems [https://www.edge.org/conversation/thomas\\_metzinger-benevolent-artificial-anti-natalism-baan](https://www.edge.org/conversation/thomas_metzinger-benevolent-artificial-anti-natalism-baan) could not only possess a much higher degree of mental autonomy and internal consistency than we do, but that they'd likely be much better at moral <https://aeon.co/essays/creating-robots-capable-of-moral-reasoning-is-like-parenting> cognition, too.

For those of us willing to face up to facts, mind-wandering creates a host of new normative issues. It challenges wide-spread background assumptions held in economics and political philosophy, all of which is profoundly relevant for applied ethics and policymaking. Climate change, predatory capitalism and religious fundamentalism really are only 'surface' phenomena, because the human actions leading to them are determined by the deep structure of our minds, the functional architecture of our culturally embedded brains. The real challenge is not climate change as such; the true problems humankind must confront are the inbuilt features of our very own minds, such as systematic rationality and empathy deficits, self-

deception, and the extreme vulnerability of the inner mechanisms creating mental autonomy.

This process of learning about ourselves is emotionally unattractive to many. It forces us to confront centuries-old systems of systematic self-deception, often represented by certain forms of organised religion. Indeed, one of the interesting aspects of mental autonomy is our general inability to notice when we *lack* it. Not only does there seem to be a widespread form of ‘introspective neglect’ related to the losses of cognitive self-control characterising our inner life, resembling a form of anosognosia or anosodiaphoria (as in patients unable to become aware of, or who remain indifferent to, the existence of a disability, such as the fact that their left arm is paralysed). The phenomenon of mind-wandering itself is also clearly related to denial and confabulation. Introspective experience and the verbal reports of one’s own inner life are probably likely to be distorted by overconfidence bias regarding the degree to which one actually is in control, by illusions of superiority and introspection illusion (in which we falsely assume direct insight into the origins of our mental states, while treating others’ introspections as unreliable).

Given all the new evidence, how can we systematically increase our own mental autonomy – and what is the role of political institutions in supporting the individual citizen in the process?

**M**editation research is poised to make major contributions to mental autonomy. Mindfulness practice can sometimes lead to a crystal-clear and silent mind that is not clouded by thoughts at all, the pure conscious experience of mental autonomy *as such* that arises without actually exerting control. In long-term practitioners, this can result from the cultivation of a kind of inner non-acting that includes noticing, gently letting go, and resting in an open, effortless state of choiceless awareness.

However, in the beginning, meditation clearly involves making decisions, as subjects develop <https://mindrxiv.org/kjyr8/> meta-awareness, alongside an awareness of their capacity for attentional control. This can be seen as a systematic form of ‘experience sampling’. A scientist in the lab might probe the frequency of mind-wandering by asking the subject to press a buzzer every time the mind wanders. In meditation, your own awareness is the buzzer: in the lab of your head, you probe your own mind, trying to detect the next dolphin as early as possible and catch yourself mind-wandering. At the beginning, the process of actively disengaging from every arising thought, again and again, for thousands of times, will create brief illusions of control as discussed above.

Whether this sort of cognition really requires a robust notion of selfhood, as most

Western philosophers would argue, would be disputed in many Eastern traditions. Here the highest level of mental autonomy is often seen as a form of impersonal witnessing or (in the words of the Indian-born philosopher Jiddu Krishnamurti) 'observing without an observer' (though even this pure form of global meta-awareness still contains the implicit knowledge that the organism *could* act if necessary). There seems to be a middle way: perhaps mental autonomy can actually be experienced as such, in a non-agentive way, as a mere capacity. The notion of 'mental autonomy' could therefore be a deep point of contact where Eastern and Western philosophy discover common conceptual ground.

## The human adult's conscious model of the 'self' might be an enculturated *post-hoc* confabulation

Empirical research will be equally important to enhancing mental autonomy. For example, it has been shown [<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4870246/>](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4870246/) that individual thoughts have variable degrees of 'stickiness', as it is more or less difficult to detach from them. This property of thoughts is related to one's mood and performance. Other recent experiments [<http://www.sciencedirect.com/science/article/pii/S105381191630074X>](http://www.sciencedirect.com/science/article/pii/S105381191630074X) with experienced mindfulness practitioners have begun to shed light on the dynamics of neural recruitment surrounding the spontaneous arising of thoughts. The difficulty is that it's very challenging for most people to accurately detect the onset of thoughts as they happen. Personally, I find that the very first one is an inevitable moment of surprise, distraction and cloudedness, and that I can at best become aware of the onset of the *second* dolphin leap. But advanced mindfulness meditators are perfect subjects, because they possess the introspective ability needed to identify the precise temporal onset, and can help neuroscientists pin down the various brain regions supporting the appearance of spontaneous thoughts.

On the other hand, if the ideas sketched in this essay are correct, then modern consciousness science might actually provide meditators with new perspectives and, perhaps, some deeper insight into their own contemplative practice. If even thoughts of the type 'I have just successfully disengaged from a thought and regained meta-awareness!' are ultimately just another illusion of control, then the practitioner has just identified with the content of a particularly sticky and clever new self-model in the brain.

It's important to remember that neuroscience isn't the only piece of the puzzle. Culture plays its part, too. This context shapes the way we report our own inner experiences, which end up consisting of inaccurate but functionally adequate

representations in our conscious minds – such as the conscious sense of moral agency, at least as reported by shallow introspectors. Accountability and ethical responsibility might actually be implemented in the human brain from above, via early social interactions between children and adults. If we tell children at an early age that they are fully responsible for their own actions, and if we accordingly punish and reward them, then this assumption will get built into their conscious self-model. They will always automatically predict that they will be held responsible, and their inner life-narrative will say ‘It has always been like that!’ since the dawn of memory. The human adult’s conscious model of the ‘self’ might therefore be an enculturated *post-hoc* confabulation, at least in part – a causal-inference illusion that’s become part of how we model our own sociocultural niche, ultimately based on how we’ve internalised social interactions and ingrained language games.

**O**n closer inspection, the mind-wandering network does not, I believe, actually produce thoughts. It also is not conscious – the person as a whole is. Rather, it creates what I would [describe <http://predictive-mind.net/papers/the-problem-of-mental-action>](http://predictive-mind.net/papers/the-problem-of-mental-action) as *cognitive affordances*, opportunities for inner action. In the theory of psychology developed by J J Gibson, what we perceive in our environment aren’t simply objects, but possible actions: this is something I could *sit* on, this is something I could *put* into my mouth, and so on. Cognitive affordances are possible *mental* actions, and they are not perceived with our sensory organs but they are available for introspection: this is a pleasant sexual fantasy I could *pursue* further, this is a potentially interesting philosophical argument I might *develop in my mind*, this is a negative feeling caused by a dim memory which I could *explore* further.

Cognitive affordances are actually precursors of thoughts, or *proto-thoughts*, that call out ‘Think me!’ or ‘Don’t miss me – I am the last of my kind!’ Our inner landscape is full of these possibilities, which we must constantly navigate. What mind-wandering does is create a fluid and highly dynamic task-domain. Every spontaneously occurring ‘task-unrelated’ mental event is a potential task in itself, a cognitive affordance, a dynamic state that has the potential to be selected and transformed from unintentional mental behaviour into something subjectively experienced as full-blown mental action.

One central function of mind-wandering, then, could be to provide us with an internal environment of competing affordances, accompanied by possible mental actions, which have the potential to become an extended process of controlling the content of your own mind. This inner landscape could even be below our conscious awareness, but it is out of this that the epistemic-agent model emerges, like any other conscious experience, seemingly selecting what she wants to know and what she wants to ignore. The open question is if it is this conscious model *itself* that plays the decisive causal role in helping the organism to predictively control its inner life. If the general

theory is correct, then mental autonomy has nothing to do with simplistic ideas such as Cartesian ‘free will’. Rather, true autonomy is about different levels of context-sensitivity and supple self-control.

Mental autonomy is not an all-or-nothing phenomenon. Clearly, it comes in degrees. Excessive social-media consumption and certain psychoactive substances damage it. Yet many mind-wandering episodes feel as if they are actually self-initiated <https://link.springer.com/article/10.1007/s00426-014-0617-x>, and there is a difference between ‘zoning out’ (losing yourself in a daydream) and ‘tuning out’ (indulging in an interesting fantasy during a boring lecture, while checking back in from time to time). What we need is an evidence-based, differentiated picture that is open for future discoveries, perhaps leading us to the insight that even the highest degree of mental autonomy is still a form of unintentional mental behaviour – without ultimate origination, but adding in veto-control, flexibility and context-sensitivity.

**G**iven what we know to date, should we eliminate the concept of ‘mental action’ altogether, and become selflessly observing Radical Neuro-Buddhists? An alternative philosophical role model would be the ‘Romantic Revisionist’. She says: ‘Now, let’s not pour the baby out with the bathwater! Let us introduce a weaker, empirically plausible notion of mental action, and then distinguish between different levels of unintentional mental events, each level characterised by its own potential for self-determination and degree of autonomy!’

With affectionate indifference, the Neuro-Buddhist calmly replies: ‘Let us be intellectually honest. In the scientific worldview, there is just no place for ultimate origination, and in one way or another that will always have to be the core of any philosophically satisfying, rich theory of what an action is. Let us not be fooled by the fairytale the conscious brain tries to tell the rest of the organism. I am willing to bite the bullet: in this world, there are only events, and no real mental actions.’

To which the Romantic Revisionist answers: ‘You just want to believe in something. You talk as if we already knew what consciousness is, and as if the scientific worldview was complete. How would you then explain the human capacity to impose rules on oneself, to actively select goals, the ability for rational guidance and, most importantly, for the intentional termination of an ongoing mental process? The predictive self-model story is just one theory among others – what really is intellectually dishonest are sweeping conclusions and your slightly complacent form of radicalism. Let’s just wait a little bit longer, the verdict is not yet in.’

**There can be no politically mature citizens without a sufficient degree of mental autonomy**

What is clear by now is that our societies lack systematic and institutionalised ways of enhancing citizens' mental autonomy. This is a neglected duty of care on the part of governments. There can be no politically mature citizens without a sufficient degree of mental autonomy, but society as a whole does not act to protect or increase it. Yet, it might be the most precious resource of all. In the end, and in the face of serious existential risks posed by environmental degradation and advanced capitalism, we must understand that citizens' collective level of mental autonomy will be the decisive factor.

Here's a positive proposal: to get started, we should aim at a productive cross-fertilisation of the two strongest aspects of the human mind. The first is our recently evolved capacity for self-critical rational thinking. At least sometimes, human beings are sensitive to rational argument. And the second is the enormous depth of our phenomenological state-space. Because of its many dimensions, the number of possible conscious states for a human being is incredibly large. We are only rarely aware of this fact, and we haven't really started to systematically test how we might deliberately alter our state-space so as to enhance our autonomy and increase experiential forms of self-knowledge, ideally backed by the rigour of modern-day neuroscience. An exception, of course, are certain ancient spiritual techniques, which also work to stabilise the neurocognitive conditions required for rationality and rigorously logical, critical thinking. Statistically, mindfulness and mind-wandering are opposing constructs, and rational thought critically depends on the capacity for attentional self-control.

It was William James, the father of American psychology, who said in 1892: 'And the faculty of voluntarily bringing back a wandering attention over and over again is the very root of judgment, character, and will. [...] And education which should improve this faculty would be *the* education *par excellence*.' We can finally see more clearly what meditation is really about: over the centuries, the main goal has always been a sustained enhancement of one's mental autonomy.

I think ideologically charged debates about freedom of the will and Stone Age reductionism are now a thing of the past. But the path forward is not about finding the right philosophical theory. It is about starting a process of actively implementing what we already know into societal and cultural practice. As a working concept, mental autonomy is an excellent new candidate for a basic value that could guide us in education, policymaking and ethics. The two-component proposal gives a new and deeper meaning to the old Kantian ideal of 'man's emergence from his self-imposed immaturity'. We might call it raising the standard of civilisation on the mental level, or developing a 'culture of consciousness'.

Finally, mental autonomy brings together the core ideas of both Eastern and Western

philosophy. It helps us see the value of both secularised spiritual practice and of rigorous, rational thought. There seem to be two complementary ways to understand the dolphins in our own mind: one, from the point of view of a truly hard-nosed, scientifically minded tourist on the prow of the boat; and two, from the perspective of the wide-open sky, silently looking down from above at the tourist and the dolphins porpoising in the ocean.

*Thomas Metzinger is full professor and director of the theoretical philosophy group and the research group on neuroethics/neurophilosophy at the department of philosophy at the Johannes Gutenberg University of Mainz in Germany. From 2014-19, he is a fellow at the Gutenberg Research College. He is the founder and director of the MIND group <<https://fias.uni-frankfurt.de/mindgroup/>>, and adjunct fellow at the Frankfurt Institute of Advanced Studies in Germany. His latest book is *The Ego Tunnel* (2009).*

aeon.co22 January, 2018