

---

# (Dis-)Attending to the Body

## Action and Self-Experience in the Active Inference Framework

Jakub Limanowski

---

Endogenous attention is crucial and beneficial for learning, selecting, and supervising actions. However, deliberately attending to action execution usually comes with costs like decreased smoothness and slower performance; it may severely impair normal functioning and, in the worst case, result in pathological behavior and self-experience. These ambiguous modulatory effects of attention to action have been examined on phenomenological, computational, and implementational levels of description. The active inference framework offers a novel and potentially unifying view on these aspects, proposing that actions are enabled by attentional modulation based on expected precision of prediction errors in a brain's hierarchical generative model. The implications of active inference fit well with empirical results, they resonate well with ideomotor action theories, and they also tentatively reflect many insights from phenomenological analysis of the "lived body". A particular strength of active inference is its hierarchical account of motor control in terms of adaptive behavior driven by the imperative to maintain the organism's states within unsurprising boundaries. Phenomena ranging from movement production by spinal reflex arcs to intentional, goal-directed action and the experience of oneself as an embodied agent are thus proposed to rely on the same mechanisms operating universally throughout the brain's hierarchical generative model. However, while the explanation of movement production and sensory attenuation in terms of low-level attentional modulation is quite elegant on the active inference view, there are some questions left open by its extension to higher levels of action control—particularly about the accompanying phenomenology. I suggest that conceptual guidance from recent accounts of phenomenal self- and world-modeling may help refine the active inference framework, leading to a better understanding of the predictive nature of embodied agentive self-experience.

### Keywords

Active inference | Action | Attention | Ideomotor theory | Intentional action | Lived body | Minimal phenomenal selfhood | Motor control | Precision-modulation | Self-model

### Acknowledgements:

I would like to thank Felix Blankenburg, Ryszard Aukstulewicz, Thomas Metzinger, and Wanja Wiese for their helpful comments.

## 1 When Attending to the Body Impairs Performance

A centipede was happy – quite!  
Until a toad in fun  
Said, "Pray, which leg moves after which?"  
This raised her doubts to such a pitch,  
She fell exhausted in the ditch  
Not knowing how to run.  
*Kathrine Craster (1871)*

The "Centipede's dilemma" nicely captures the fact that I am usually not paying attention to my body as I interact with the world. The poem also suggests that this may be a good thing, as such attention—triggered, for example, when one is asked how one coordinates one's many legs—can severely impair one's normal functioning in the world: the centipede certainly *wants* to move, yet it fails to do so because it directs its attention towards its body (i.e., to how the body should execute movements) instead of forgetting about it and just moving as usual. The suppression of the body from experience has been of central interest to classical phenomenology, which considers it a necessary means for

interacting with the world as a “lived body”. The lived body concept was proposed by Merleau-Ponty (Merleau-Ponty 1945/1962, and was developed by others, see Gallagher 1986, for a review) to explain, without resorting to Cartesian dualism, the dual role of the body as both an object belonging to the world, and our means (the “vehicle”) of being an experiencing and acting subject in this world. In brief, the lived body *is* our being and acting in the world; it therefore is a “lived body-environment” (Gallagher 1986, p. 162). In such equilibrium with the environment, the body is not an object in my phenomenological field—it is absent from my experience.<sup>1</sup> Naturally, the body can be experienced via the senses—but this explicit, often “analytic” access to the objective body reveals “its belongingness to the physical realm” (Legrand 2011, p. 15; cf. Merleau-Ponty 1945/1962; Liang 2015). Husserl, the founder of phenomenology, described this as a “self-objectivation of the lived body” (Zahavi 1994, p. 70). While this need not necessarily imply a total loss of the body’s subjectivity (cf. Zahavi 1994), it may lead to an experienced “doubling of the body, the ‘splitting of the phenomenon’ into two abstractions” (Gallagher 1986, p. 140).

An important postulate of classical phenomenology of the body is that “it is never our objective body that we move, but our phenomenal body” (Merleau-Ponty 1945/1962, p. 106). If we subscribe to this postulate, we can see why directing attention to the body may be detrimental to (inter)acting in the world: self-directed attention presents the body *also* as an object of experience, thus interfering with the normal experience and performance as a lived body-environment under experiential suppression of the physical body. This can happen in two ways: the body can suddenly appear as an object in my phenomenological field, such as when I bump into something, when I am exhausted, or when I am injured (i.e., in “limit situations”, Gallagher 1986, p. 148). In these cases, the body-as-object captures my attention. But similar self-objectivation can also be induced deliberately via endogenous self-directed attention, as in the case of the centipede. An extreme example of this is illustrated by the “analytical, decomposing effect” (Fuchs 2010, p. 241) of self-directed attention in schizophrenic hyperreflexivity, where “every action, however trifling, requires targeted attention and action of the will, as it were, a ‘Cartesian’ impact of the Ego on the body” (p. 247) and “the self is, so to speak, no longer at home in its body” (p. 251). Thus some forms of mental illness may be understood as an extreme case of experiencing the body as an object, which may result in a vicious cycle<sup>2</sup> of increasing “estrangement from oneself” (Fuchs 2010, p. 239)—estrangement from oneself as a lived body. Less extreme, but similar cases include directing attention towards automatic behavior that has not been learned, like falling asleep or being sexually aroused (Fuchs 2010). Such an impairment of performance by self-directed attention also underlies numerous reports of professional athletes who suddenly become unable to perform certain long-mastered movements. A prominent case is former baseball pitcher Steve Blass, who had to quit his career after suddenly and inexplicably losing his ability to throw accurately. Presumably, just like the centipede such athletes start focusing too much on the movement execution itself.

Of course, how attention affects action has also long been a central empirical research question. Research on motor control has demonstrated that endogenous attention is essential for learning, selecting, and supervising actions. However, experiments have also shown that deliberate attention to action execution usually comes with costs like lack of smoothness and slower, step-by-step performance (Norman and Shallice 1986; Diedrichsen and Kornysheva 2015). For example, people perform and learn motor tasks worse when they attend to their execution, whereas performance increases and

1 This has nothing to do with the awareness that one actually *is* a physical body per se, as this fact can be implicitly experienced without directing attention to it, just as I can be aware that I am walking without directing attention to my walking (Merleau-Ponty 1945/1962; Norman and Shallice 1986; Gallagher 1986). There have been some attempts to clarify why being aware of oneself as a physical body does not necessarily imply a suspension of the body’s subjectivity (cf. Zahavi 1994). For example, Legrand (Legrand 2011) proposes a distinction between analytic and subjective access to the self-as-object, where only the former implies a disruption of the body’s subjectivity by reification. Alternatively, Liang (Liang 2015) distinguishes the first-personal from the third-personal sense of body ownership, where the latter treats the body as an object.

2 There may be a relation to sustained attention directed from the meaning to the carrier of that meaning, as in the case of semantic satiation, where continued fixation or verbal repetition of a word causes the word to lose meaning (Fuchs 2010; cf. Hohwy 2012; Clark 2015).

movement is much smoother when attention is directed away from execution (e.g. [Wulf et al. 2001](#)). Detrimental effects are particularly evident when attention is deliberately directed to already well specified (learned) movements (which may be described as a “reinvestment in movement”, [Brown et al. 2013](#), p. 421), where such an internal focus of attention may enslave resources and interfere with automatic motor control processes or schemata ([Wulf et al. 2001](#)).

In sum, the ambiguous modulatory effects of attention to action—a necessary control mechanism on the one hand, a potentially substantial impairment on the other—have been examined from various perspectives, spanning phenomenological, computational, and implementational levels of description. In the remainder of this paper, I will argue that all of these levels can in principle be accommodated by the active inference framework ([Friston et al. 2009](#)), a recent mechanistic account of adaptive behavior as being driven by hierarchical prediction error minimization which is ultimately aimed at occupying unsurprising states, and which appeals to a theory of brain function based on a universal free energy principle (FEP, [Friston 2010](#); cf. [Hohwy 2013](#); [Clark 2015](#)). I will first present an explanation of the aforementioned ambiguous effects of attention to movement in terms of active inference, i.e., as attentional modulation at low levels of the motor control hierarchy of the central nervous system. I will then examine the claim that active inference can in principle be extended to *all* levels of action and behavior—thus mapping, for instance, onto concepts like intention and cognitive control. I will argue that the active inference framework may help bridge the various levels at which attention to the own moving body has been investigated, and thus constitutes a very promising basis for an interdisciplinary investigation of embodied agentic self-experience.

## 2 Active Inference: Moving by Attentional Modulation

The FEP is built around the claim that biological agents must maintain homeostasis and must therefore occupy a limited range of states defined by their phenotype. Thus avoiding “surprising” states is the common principle underlying all behavior and cognition ([Friston et al. 2009](#); [Friston et al. 2010](#); [Friston 2010](#)). However, the state of the environment (including the organism itself) is hidden from the agent and must be inferred from incoming sensory information. The FEP proposes that the brain performs such inference via probabilistically mapping hidden causes to sensory data in a hierarchical generative model (HGM), where each level encodes conditional expectations (“beliefs”) about information in the level below, with the overall hierarchy ultimately modeling the generative process in the environment that causes the current sensory data. By inverting this model, surprise approximated in the form of prediction error<sup>3</sup> can be minimized via model (parameter) update, which is known as *predictive coding* ([Friston et al. 2009](#)): ascending data (at the lowest level, actual sensory input) are compared with descending predictions at each level, and only unpredicted data—the prediction errors—are communicated upwards. These errors can be then minimized by changing the model’s higher-level beliefs about the causes of this input, which corresponds to perceptual inference ([Friston 2010](#)). Predictive coding in the brain therefore emerges as a consequence of the imperative to maintain homeostasis, whereby priors may be acquired and optimized by learning, or be innate and optimized by natural selection ([Friston et al. 2010](#); [Pezzulo et al. 2015](#)).

For inference to be optimal, the brain needs to decide which prediction errors are currently most relevant, and it needs to assign these errors relatively more weight in determining inference. According to predictive coding, this is implemented by adjusting the gain of prediction error units according to their expected precision (which corresponds to reliability or inverse variance). Thus there are two types of descending predictions: those of input, inhibiting error units in the level below, and those of precision, optimizing the gain of error units (i.e., changing the postsynaptic response of error units to their presynaptic inputs, presumably via NMDA-dependent plasticity, dopaminergic modulation,

<sup>3</sup> “Surprise” corresponds to the negative log-likelihood of the sensory data under the model and can be approximated via free energy, which, under some simplifying assumptions made by the predictive coding scheme, corresponds to prediction error ([Friston et al. 2010](#)).

or other classical neuromodulators; Friston et al. 2012a; Adams et al. 2013). Precision-modulation is thus also a Bayes-optimal mechanism that minimizes free energy. Weighting prediction error signals by their expected precision determines their relative impact on inference, i.e., on updating prior beliefs at higher levels of the model (Friston et al. 2009). Applied throughout a HGM's hierarchy, this mechanism allows for delicately balancing the relative influence of sensory evidence and prior beliefs on (active) inference. This top-down modulation is a contextual one that will vary depending on the current circumstances and requirements: "When higher levels have greater precision, their contextual influence dominates; whereas, when expected sensory precision is high, inference and subsequent behavior is driven by sensory evidence" (Pezzulo et al. 2015, p. 24). Note that the top-down, context-dependent selection and weighting of (sensory) prediction errors, based on their expected precision, is nothing other than weighting specific sensory channels according to (expected) signal-to-noise ratios, the function generally attributed to *attention* (Feldman and Friston 2010; Aukstulewicz and Friston 2016). Under active inference, precision-modulation is therefore described as an attentional modulation (e.g. Edwards et al. 2012; Brown et al. 2013). Crucially, such attentional modulation also determines whether an agent resorts to perceptual inference as described above—or whether it acts.

Of course, not only can we suppress prediction error by changing our model so that it better reflects the state of the world, we can also change the state of the world so that our sensory input corresponds to our current predictions. By acting on the environment, i.e., by intervening with the generative process itself, we can directly suppress surprise (i.e., free energy) and thus also minimize prediction error. *Active inference* (Friston et al. 2009) thus extends the principles of predictive coding (as described above for sensory systems) to the motor system—the difference is that in the motor system, the predictions and errors are proprioceptive, i.e., they are about the posture and position of the body's joints and the forces applied to them (Adams et al. 2013; Friston et al. 2012a). Active inference therefore explains motor control in terms of predicting states of the body as part of the "environment", i.e., of the hidden process generating the current sensory data.<sup>4</sup> Movement accordingly occurs because high-level multimodal or amodal beliefs predict counterfactual exteroceptive and proprioceptive states (sensory consequences that would ensue if the movement were performed), and the proprioceptive predictions generate a prediction error in the spinal cord where they meet afferent information about the current proprioceptive state, i.e., movement is predicted but not sensed. Unlike sensory systems, where the predictions would now be revised to explain away prediction error, the motor system can use an alternative strategy to suppress errors: the fulfillment of proprioceptive predictions by activation of alpha motor neurons of classical reflex arcs in the spinal cord, i.e., by performing the predicted movement (Friston et al. 2010; Friston et al. 2011; Adams et al. 2013; Brown et al. 2013; Edwards et al. 2012). Thus movement results from predictions about its sensory consequences rather than from motor commands in the classical sense.

Under active inference, goal states for action and behavior are defined by prior expectations. Motivated or adaptive behavior can therefore be described as based on the minimization of interoceptive prediction error (which informs about deviance from optimal homeostatic levels) and proprioceptive and exteroceptive prediction error (which specifies the external goal state to be attained by action, Pezzulo et al. 2015). An important implication of this, one which distinguishes active inference from previous approaches, is that goal states are not desirable because they are "valuable" in themselves, but because they are states that the organism *expects* to occupy (under the assumption that it will always minimize free energy). In other words, we selectively sample sensory input that is expected, based on the predictions of our current HGM, to be precise—thus action and perception are intimately coupled (Friston et al. 2009; Friston et al. 2011; Friston et al. 2012b; Hohwy 2012; Pezzulo et al. 2015).

4 One could say, more specifically, that defining the environment in this sense includes the physical body but not the brain that actually employs the HGM, as the brain's states are (to our knowledge) not accessible to itself via sensory organs. Some interesting questions that follow from this are discussed by Metzinger (Metzinger 2017).

Crucially, whether or not action occurs is determined by an attentional balancing act, i.e., attention weights prediction errors to optimize not only perceptual inference, but also action. Movement only occurs if the proprioceptive prediction errors at the spinal cord level generated by confident high-level sensorimotor expectations are expected to be very precise, and if simultaneously the expected precision of ascending sensory prediction error (which conveys evidence against the prediction that one is moving) is attenuated. Only then are proprioceptive prediction errors at the spinal level acted out instead of being accommodated by perceptual inference, i.e., only then are the counterfactual predictions about the body's state fulfilled rather than updated (Friston et al. 2009; Friston et al. 2011). Thus, in addition to confident beliefs about the sensory consequences of the intended movement, “action requires [...] targeted dis-attention” away from current sensory evidence that one is actually not moving (Clark 2015, p. 217).

Within this framework, the detrimental effects of attention towards movement execution—remember the centipede—are readily explained in terms of low-level precision-modulation (Brown et al. 2013; Edwards et al. 2012): attending to the sensory input generated by my body increases the precision of the corresponding sensory prediction errors, which are conveying evidence contra the descending predictions of the sensory consequences generated by movement. These errors have now more influence on higher-level beliefs, which are therefore adjusted to accommodate the fact that I sense no movement. Consequently, no sufficiently precise proprioceptive prediction errors are generated, and no (or abnormal) movement results (Adams et al. 2013). Therefore, a system following active inference is only capable of producing movement under an appropriate balance between precision at high versus low levels; under abnormal precision-estimation, pathological behavior ensues, with effects varying according to the hierarchical site of the imbalance (Brown et al. 2013; cf. Edwards et al. 2012; Friston et al. 2012a).

In sum, the act of balancing expected precision at various levels of the generative model determines whether a system operating on such a model resorts to perceptual inference or to action. An important (and prima facie counterintuitive) implication of active inference is that such attentional control is not an action but a part of perceptual inference—it is optimization of precision in a HGM that “has no notion of action; it just produces predictions that action tries to fulfil” (Friston et al. 2009, p.4). Action or behavior—a change of external states—emerges only at the lowest level of the motor hierarchy as a suppression of precise proprioceptive prediction error by peripheral neurons (the central nervous system is only concerned with perceptual inference) and an attenuation of the expected precision of ascending sensory prediction error. Describing the underlying precision-modulation as (endogenous) attentional modulation implies the Jamesian characterization of attention as something selective that “implies withdrawal from some things in order to deal effectively with others” (James 1890, p. 404). Specifically, to be able to interact with the world, I need to withdraw attention from my body's current state and focus it on what I predict sensing in my desired state. This conclusion is very similar to that of classical phenomenology, namely, that the “experiential absence” of the body is necessary for action in the world—being a lived body-environment—and that attention directed towards the objective body is detrimental to normal performance. So active inference intuitively explains why the centipede cannot move in terms of specific effects of low-level attentional (precision) modulation. But does this mechanism likewise explain why the centipede can normally move as it wishes?

### 3 Attentional Modulation throughout the Hierarchical Generative Model: A Motor Control Hierarchy

One of the greatest strengths and boldest claims of the active inference framework is its proposed universal mechanism operating across all levels of the HGM, which is neurobiologically implemented via predictive coding in the brain (and action via reflex arcs). It acknowledges the hierarchical nature of motor control which spans from kinematics to conceptual knowledge about the world; it integrates

distinct control systems (Friston 2011; Pezzulo et al. 2015), and it avoids the pitfalls of describing action control either as purely stimulus-driven, or in purely “perceptuo-motor” or “associative” terms (Ondobaka and Bekkering 2012; cf. Kilner et al. 2007). Active inference thereby fundamentally relies on the top-down contextualizing effect of higher levels on lower ones, enabled by attentional modulation based on expected precision, where a context can be a selected action, a goal, or even agency. Thus it aims at explaining phenomena across all levels of the motor hierarchy, from the reflex arcs that produce movement to intentional action and cognitive control (Pezzulo and Cisek 2016).

### 3.1 Sensory Attenuation and Agency

A particularly interesting implication of the active inference account is its explanation of sensory attenuation, which can be observed during movement (Blakemore et al. 1998; Brown et al. 2013) and even during movement preparation (Voss et al. 2006). The attenuation of self-generated sensory signals during movement has previously been proposed in terms of forward models that predict and thus cancel out the sensory consequences of one’s movements based on the body’s current state and corollary discharge (Blakemore et al. 1998; cf. Friston et al. 2012b for a more detailed comparison of these accounts). However, the implications of attentional balancing across the motor hierarchy as assumed by active inference go beyond this: as noted above, sensory attenuation is a necessary dis-attention away from sensory input, which would otherwise bias perceptual inference and potentially preclude movement (as is likely the case in the centipede’s dilemma). Active inference even postulates that sensory attenuation and its effect on perceptual inference underlies certain forms of self-consciousness, including the experience of self-other distinction in action execution versus observation. Similarly to previous accounts of the mirror neuron system, active inference assumes that the brain uses the same HGM and thus the same action control hierarchy to model and predict the intentions, goals, actions, and kinematics of both one’s own and other bodies (Kilner et al. 2007). This means that high-level beliefs encoding action goals “do not assign agency to any particular agent” (Friston et al. 2012b, p. 539): these beliefs generate amodal, multimodal, and unimodal predictions throughout the motor hierarchy for one’s own and for others’ actions.

According to active inference, self- or other-agency—whether I perform a movement or whether I instead perceive someone else performing the movement—is a *context* determined by precision-modulation of (i.e., selective attention to) proprioceptive and visual information in one and the same HGM (Friston et al. 2011). If I observe an action, the visual prediction error generated by the seen movement will update multimodal beliefs at higher levels in the motor hierarchy, which predict visual and proprioceptive action consequences. This means I must attenuate the expected precision of the prediction errors generated by *proprioceptive* predictions—otherwise I might move myself. With this attenuation, updating my model’s beliefs by visual prediction errors allows me to infer the cause of the observed movement and thus ultimately to understand the other’s intentions (Kilner et al. 2007; Friston and Frith 2015). Conversely, recall that increased high-level proprioceptive precision is necessary to produce movement via spinal reflex arcs. Thus “active inference presents in one of two modes; either attending to sensations or acting during periods of sensory attenuation” (Friston and Frith 2015, p. 398), where attentional modulation is fundamentally involved in realizing both of these modes.<sup>5</sup>

Correspondingly, misattributions of agency, as in schizophrenia, have been explained by aberrant attention. Here, inference about the hidden causes of sensations fails because the precision of high-level beliefs is (falsely) increased to compensate for a failure to attenuate sensory prediction error during action. These overconfident beliefs generate additional, incorrectly confident, predictions about external causes—the agent is not able to infer whether it caused its sensations itself, or whether someone or something else caused them (Brown et al. 2013). In sum, under active inference, agency is grounded

<sup>5</sup> However, the *sense* of agency in action certainly also depends on behaving in accordance with (confidently) expected states, i.e., when precise proprioceptive prediction error is resolved in line with our predictions (Hohwy 2007; Hohwy 2013; Friston et al. 2013; Clark 2015).

in the contextual influence of high-level beliefs on lower levels, which manifests itself in attentional modulation, i.e., in adjusting the relative gain of vision and proprioception.

### 3.2 Intentional Action as Adaptive Behavior

So far, we have seen that active inference provides an elegant explanation for the role of precision-modulation (attentional biasing) in movement initiation and production. However, active inference aims to explain all facets of behavior. Therefore even complex phenomena like the conscious selection of actions based on goals and intentions should be explained as driven by beliefs about behavior and modulated by expected precision. The proposed answer that active inference offers to these questions is partly reminiscent of that of the classical ideomotor theory (IMT) of action, which was developed as an explanation for how intentions might drive actions (James 1890; Stock and Stock 2004; Kunde et al. 2007). Here, I will briefly outline some commonalities and differences between IMT and active inference, which will reveal the novel contribution and explanatory power of active inference, but also some questions that it leaves open.

Most people would probably agree that an intentional action is always accompanied by a conscious goal representation (cf. Hommel 2015). IMT<sup>6</sup> proposes that this conscious goal representation is in fact driving the action. Movement is accordingly brought about by an “idea” or “effect image” of the anticipated sensory consequences of that movement, which is itself the result of previous associative learning between movements and their sensory consequences (Hommel et al. 2001; cf. Stock and Stock 2004, for a review). Consequently, IMT states that, rather than there being separate perceptual representations and motor commands, perception and action share a common representational format (Prinz 1997; Hommel et al. 2001), just as the active inference view does not distinguish between perceptual and motor representations in the classical sense. An interesting conclusion of IMT is that even the simplest actions are goal-directed, as they are always aimed at reaching an anticipated sensory effect (the “goal representation”, Kunde et al. 2007; Hommel 2015). The same holds for active inference, where goal representations are the result of perceptual inference and correspond to (counterfactual) beliefs about sensory states that elicit corresponding prediction errors. Like active inference, IMT emphasizes that actions can only be brought about by ideas if one ignores “competing” ideas—most notably, the fact that one is currently not moving (James 1890; Clark 2015). In sum, both IMT and active inference state that a withdrawal of attention from movement execution and a focus onto the action goal is essential for action, thus nicely explaining the Centipede’s dilemma.<sup>7</sup>

Active inference, however, specifies its claim that movement relies on both confident beliefs and attenuated sensory input by suggesting an underlying attentional modulation, implemented by increasing high-level precision and decreasing low-level precision. The universal role of attention proposed by active inference, however, seems at odds with some extensions of IMT. For example, Hommel et al. 2001 differentiate between attentional and intentional weighting in perception and action:

With reference to perception, feature weighting may be called an *attentional* process, inasmuch as it selectively prepares the cognitive system for the differential processing of relevant (i.e., to-be-attended) and irrelevant (i.e., to-be-ignored) features of an anticipated perceptual event. With reference to action planning, however, the same kind of feature weighting could rather be called an *intentional* process, because it reflects the perceiver/actor’s intention to bring about a selected aspect of the to-be-produced event. (Hommel et al. 2001, p. 864)

<sup>6</sup> There are of course many variations of IMT; here I will only present the basic assumptions of its classical form.

<sup>7</sup> Experimental work has shown that even visual space is attentionally structured in this way: whereas attentional processing is facilitated in peri-hand space, it is impaired on the hand’s surface (Taylor and Witt 2014). The explanation may be the same: the brain prevents attention to the body to assist goal-directed action.

Active inference, in contrast, specifies the intentional process as *attention to intention* (Edwards et al. 2012), where an intention is specified by a high-level goal representation. In fact, attention to action intention increases brain activity in supplementary motor areas that under active inference encode intentions (Lau et al. 2004). Active inference thus subscribes to James' proposal that "attention *creates* no idea; an idea must already be there before we can attend to it" (James 1890, p. 450). However, it puts attention (i.e., precision-modulation) at the center of intentional action selection. In conclusion, under active inference, goal states and intentions are *defined* by high-level beliefs, and *selected* by attention (i.e., certain beliefs are assigned more precision, cf. Friston et al. 2011; Pezzulo and Cisek 2016).

In this light, I tentatively propose, precision-optimization at higher levels of the HGM maps onto concepts like "will" (defined as "the direction of action by direct conscious control through the supervisory attentional mechanism", Norman and Shallice 1986, p. 24) or "cognitive control" (defined as the "ability to guide one's behavior in line with internal goals", Jiang et al. 2014, p. 31). In fact, active inference's tenet that attentional allocation is based on predictions of precision is similar to the proposals of some Bayesian accounts of cognitive control, where "the regulation of cognitive control should be considered as a process of *predicting* the optimal amount of cognitive control required in a given context" (Jiang et al. 2014, p. 35). Intuitively, concepts like will or cognitive control imply an important function of attention in directing intentional behavior in line with one's goals—sometimes, for example under distraction or uncertainty, such direction of actions will be notably harder. Classical theories of the relationship between attention and action have correspondingly suggested that "will varies along a quantitative dimension corresponding to the amount of activation or inhibition required from the supervisory attentional mechanisms" (Norman and Shallice 1986, p. 24). Conversely, put in the vocabulary of IMT, in situations where there is no competing "idea", there is also no need for "will" (James 1890).<sup>8</sup> Active inference likewise proposes that high-level precision is especially important under "cognitive conflict", for example, in situations where multiple representations have high precision (e.g., at high and low levels simultaneously, Pezzulo et al. 2015). In such cases, the brain needs to weight a certain belief (goal representation) more strongly than other beliefs and/or more strongly than sensory evidence. The voluntary allocation of attention against the "resistance" of some other precise belief or sensory evidence could explain why we experience an accompanying sense of effort in these situations (Metzinger 2017).

So on the one hand, conscious experience (of will and effort) could co-vary with computational cost of attentional allocation. On the other hand, however, conscious experience and top-down attentional control need not always correspond: in certain functional motor symptoms, for example, movements are executed but feel involuntary. Active inference accounts of such pathological behavior (Edwards et al. 2012) explain it as resulting from the generation of abnormally confident intermediate-level beliefs. These beliefs are sufficiently high-level to generate complex movements, but are still below the levels associated with representing the intention to move. Thus movements are induced, however, are not inferred to have been intended because the resulting percepts are not predicted by higher levels. Hence, although these movements are produced by voluntary (top-down) attention, they do not feel voluntary. This explanation aligns with previous observations that there are "cases in which one experiential sense of 'automatic' does not correspond to 'automatic' in the operational sense" (Norman and Shallice 1986, p. 19), and so some action may seem automatic while actually involving volitional attentional top-down control. For the centipede, the reverse case seems to be true: it does not *want* to be immobile—it wants to move!—but its voluntary attentional allocation prevents this.

Like any other account of action, active inference now faces the challenge to explain which aspects of motor control are accessible to conscious experience, and why. Recent extensions of the IMT have, for example, dropped the assumption that the action-driving ideas or goal representations must be conscious, and do not consider conscious experience to play a causal role in action control (Hommel et

<sup>8</sup> James explicitly distinguished between "ideo-motor" and "willed" acts (cf. Norman and Shallice 1986).

al. 2001; Prinz 1997). Their conclusion is that voluntary action may well be possible without conscious experience (Hommel 2015). Active inference offers a convincing mechanistic theory of attention as precision-optimization during perceptual inference and action. Early accounts linking attention and motor control suggested that “the phenomenology of attention can be understood through a theory of mechanisms” (Norman and Shallice 1986, p. 25). However, while attentional modulation as part of active inference very elegantly explains low-level phenomena like sensory attenuation, its extension to higher-level phenomena such as intentional action does not (yet) immediately accommodate the *phenomenology* of attentional allocation in action control. Therefore, some explanatory work remains to be done if active inference is to fully explain all aspects of agentive self-experience: which aspects of volitional behavior are accessible to consciousness, and how does the phenomenology associated with, for example, attentional agency and conscious volition emerge from the proposed brain mechanisms? As one starting point, a valuable contribution, in the form of conceptual guidance, can come from analytical approaches to phenomenal self- and world-modeling.

#### 4 Active Inference and Phenomenal Self-Modeling

One such candidate complementary account is self-model theory (SMT, Metzinger 2004; Metzinger 2009). SMT is based on the assumption that the experience of being a self emerges in organisms or systems because they possess an internal model of the world that includes and is centered on the organism itself, which, through identification of the model with its content, experiences phenomenal selfhood (Blanke and Metzinger 2009). Such a model is therefore called a phenomenal self-model (PSM, Metzinger 2004; Metzinger 2009). There are striking commonalities between the assumptions of SMT and active inference (Limanowski and Blankenburg 2013; Hohwy 2013; Metzinger 2014). Most notably, SMT suggests a hierarchy of phenomenal self-modeling, ranging from pre-reflective, “minimal” self-representations like a first-person perspective, body self-identification, or spatiotemporal self-location (Blanke and Metzinger 2009) to complex cognitive self-representations (Metzinger 2017). Such self-modeling can be well described in terms of active inference, whereby the “self” (in all its cognitive-to-minimal dimensions) is a sophisticated hypothesis about the organism’s environment which is generated by the brain’s HGM, and which tries to maximize evidence for its own existence (Limanowski and Blankenburg 2013).

The SMT account also proposes an important universal “second-order” function of attention operating on a PSM, but a slightly different one than on the active inference view: the attentional absence or inaccessibility of certain processing stages of self-modeling determines the phenomenal transparency of the respective conscious mental representations<sup>9</sup> and the associated experience of presence or realness (Metzinger 2004). Thus mental representations are transparent because only their content, and not their vehicle (e.g., the brain processes at earlier stages underlying this representation) is accessed by attentional introspection. However, not all mental representations are fully transparent. Rather, they can become more or less transparent: the more a system in possession of a PSM can attentionally access earlier processing stages, the less transparent (or more opaque) the representation becomes. This means the representation is recognized *as* modeled: as an internal, self-generated and mind-dependent construct, rather than as an invariant property of the world (Metzinger 2004; Metzinger 2009). Transparency is thus also a “phenomenal signature of epistemic reliability” (Metzinger 2014, p. 124; cf. Seth 2015), i.e., it is a sign of the system’s certainty that it has identified something that is real. Conversely, if parts of one’s PSM become opaque, this indicates the need to question their realness, and a possible need to revise one’s self-model.

Importantly, the SMT thereby assumes a “gradient of realness in the human self-model, with the bodily self being perceived as real and present while the cognitive self-model is experienced as comprised of representations” (Metzinger 2014, p. 123). So whereas I am (or can be) attentionally aware

<sup>9</sup> This does not apply to unconscious representations; in the following, only conscious representations are referred to.

that my conception of myself as an industrious person is actually made (up) by my mind, I usually do not conceive of minimal aspects of myself as an embodied self in this way—these representations are in this sense transparent to me. In other words, while I can easily change some cognitive conceptions of myself, changes to pre-reflective representations at lower levels of my PSM like body self-identification are far more difficult to make, and I believe they may have far more severe consequences. Note that although it is certainly possible to update such lower levels of bodily self-representation, as for example one’s perceived arm position in the rubber hand illusion, this does not imply that the *realness* of the content of the underlying (still transparent) self-representation is questioned—even in the rubber hand illusion, the assumption of body-self identification holds: I still feel like a normal body with just one, not two right arms (Limanowski 2014; Hohwy 2013). However, I would speculate that even minimal self-representations, i.e., those aspects of minimal phenomenal selfhood eventually constituting the basic, bodily-founded self-experience (Blanke and Metzinger 2009) can (partly) lose their transparency. I further think that such a loss of transparency at low levels of the PSM may result in (usually temporary and reversible) pathological experience—in the worst case, I would cease to “be” a self (Metzinger 2004).

In this way, SMT offers another conception of the detrimental effects of self-directed attention onto the bodily foundations of being a self (a lived body). The SMT view proposes that transparent self-modeling gives us the feeling of “being there” in the world (Blanke and Metzinger 2009; Metzinger 2004; Seth et al. 2011; Limanowski and Blankenburg 2013; Limanowski 2014). Hence, speculatively, the phenomenal absence of the body-as-object can in SMT be conceived of as a form of transparency of the representations underlying minimal phenomenal selfhood. Conversely, once I attend to the body, these phenomenal representations may gradually (but will not necessarily<sup>10</sup>) become opaque—one could rephrase this in classical phenomenological terms as a partial loss of the body’s experiential absence. The result could be a state in which the system experiences a certain representation of the self as opaque, because it recognizes, for example, that the link between itself and this particular physical object that is the body is actually “made up” by itself. Following Metzinger’s theory, the phenomenological prediction for such states is an experience of “de-identification” as occurring, for example, in depersonalization disorder (Metzinger 2004; Metzinger 2009). A similar experience is also tentatively suggested by the phenomenological description of schizophrenic hyperreflexivity as an abnormal “reification” of the embodied self by self-directed attention, whereby “the transparency of the bodily medium gets lost” (Fuchs 2010, p. 242). Under such lost transparency of the lived body, “aspects of oneself are experienced as akin to external objects” (Sass and Parnas 2003, p. 427). Perhaps one could say that under pathological self-directed attention the body becomes more present and “real” as a physical object, just as pain becomes more present when attended to. Less dramatically, this could apply to cases of attention to movement execution, where transparent, or perhaps even unconscious representations become conscious and (partly) opaque due to endogenous attention, which interferes with normal, fluent performance.

However, what SMT can most notably contribute to active inference is an analysis of the phenomenology accompanying various levels of action control. For instance, the transparency gradient assumed in the PSM may help us understand why some of active inference’s proposed attentional modulations are very intuitive (e.g. it is very intuitive that attending to the action goal is necessary to act; goal representations are high-level, and may even be opaque, cf. Gallese and Metzinger 2003), whereas others are not so easy to grasp (e.g. low-level sensory attenuation: do I really volitionally ignore my body during movement initiation?). This could also help explaining why, although attentional modulation is functionally the same across various levels of the HGM, there is *phenomenologically* a substantial difference between whether I attend to the external world or to my bodily self (Metzinger

<sup>10</sup> Of course, SMT also entails the classical notion of attention as sharpening the representation of what is currently relevant; attending to a sensation can potentially increase transparency and also the “realness” of the resulting percept. However, in this case attention is directed to the content of the representation, not to the fact *that* the sensation is the content of a representation implemented in the brain (Metzinger 2004).

2017). SMT likewise tells us why agents act as if there are desirable goals in the world where there are really just goal representations in the agent's HGM: via transparent phenomenal modeling, the agent arrives at the conclusion “that goals, actions, and intending selves actually belong to the basic constituents of the world that it is internally modeling” (Gallese and Metzinger 2003, p. 366; however, certain goal representations can also be opaque). Recall that according to active inference, the self-other agency distinction relies on a contextual manipulation of the influence of complex higher-level beliefs on visual and proprioceptive modalities. By assuming that such high-level representations as well as the representation of the attentional allocation process itself may be transparent, SMT provides an explanation of why a system operating via hierarchical inference *experiences* itself as an agent—or conversely, why it experiences another agent as the cause of its current sensory data. Thus conscious volition emerges when an agent integrates a goal representation as an object within a “model of the phenomenal intentionality relation”, a representation of an asymmetric subject-object relationship, i.e., of “a system being *directed* at a goal state” (Metzinger 2017; Gallese and Metzinger 2003; Metzinger 2004). Attentional agency, on the other hand, is a fully transparent representation “of the process of selecting the object component for attention” (Gallese and Metzinger 2003, p. 374); it is the experience that results from identification of the agent as a whole with a particular self-representation as an “epistemic agent” (Metzinger 2017).

In sum, SMT accommodates many overlaps between the active inference framework and phenomenological analysis of bodily self-experience, and with its conceptualization of phenomenal transparency versus opacity of conscious mental representations offers a compelling complement. Therefore SMT also opens up alternative ways of addressing some open questions within the active inference framework. Active inference, conversely, offers a neurobiologically plausible implementation of hierarchical self- and world-modeling, including specific testable hypotheses about recurrent message-passing and precision-modulation in the brain. Phenomenological questions have already been addressed using this approach, for example, explaining the loss of a sense of presence due to imprecise interoceptive prediction errors (Seth et al. 2011; Seth 2013; see also Limanowski 2014; Liang 2015 for related discussions of phenomenological implications of experimental paradigms that rely on the direction of attention to specific features of the bodily self). A joint effort of active inference and SMT could be extremely useful in understanding how and why certain aspects of volitional action are conscious, and in the long run, understanding the embodied agentic self-experience in general.

## 5 Conclusion

Most of us will have experienced beneficial and detrimental effects of attention to action to some degree. Not surprisingly, the role of attentional modulation in action control—more generally, in the experience of being an embodied agent in the world—has attracted the interest of philosophers, phenomenologists, psychologists, and neuroscientists alike. This interest has resulted in many hypotheses being proposed, but has also opened up many questions. Active inference, as implemented in the brain via predictive coding, offers a very elegant mechanistic- and implementational-level explanation of adaptive behavior—ultimately, as the result of a system trying to maintain its states within unsurprising boundaries. Active inference describes what happens in the brain of the centipede when it, despite wanting to move, cannot, due to increased attention to sensory prediction errors that preclude fluent movement generation. Thereby active inference proposes attention as a mechanism that balances between the relative impact of prior beliefs and current sensory evidence on inference, thus explaining a range of empirical and phenomenological observations of both normal and pathological behavior. This explanation acknowledges the fundamental role of the body for being an agent in the world, while also emphasizing the body as being part of the to-be-predicted environment. This is very much in line with classical phenomenology's interpretation of the experiential absence of the body-as-object in the subjectively lived body-environment. The extension of the active inference account to higher levels

of action control, however, leaves open some questions about the accompanying agentic self-experience, i.e., the phenomenology of, for instance, volition or attentional agency. Here, a joint application of active inference-based views and analytical accounts of phenomenal self- and world-modeling can lead to conceptual refinement and a correspondingly enhanced understanding of the predictive nature of action and self-experience.

## References

- Adams, R. A., Shipp, S. & Friston, K. J. (2013). Predictions not commands: Active inference in the motor system. *Brain Structure and Function*, 218 (3), 611–643.
- Auksztulewicz, R. & Friston, K. (2016). Repetition suppression and its contextual determinants in predictive coding. *Cortex*, 80, 125–140.
- Blakemore, S.-J., Wolpert, D. M. & Frith, C. D. (1998). Central cancellation of self-produced tickle sensation. *Nature Neuroscience*, 1 (7), 635–640.
- Blanke, O. & Metzinger, T. (2009). Full-body illusions and minimal phenomenal selfhood. *Trends in Cognitive Sciences*, 13 (1), 7–13.
- Brown, H., Adams, R. A., Parees, I., Edwards, M. & Friston, K. (2013). Active inference, sensory attenuation and illusions. *Cognitive Processing*, 14 (4), 411–427.
- Clark, A. (2015). *Surfing uncertainty: Prediction, action, and the embodied mind*. New York: Oxford University Press.
- Diedrichsen, J. & Kornysheva, K. (2015). Motor skill learning between selection and execution. *Trends in Cognitive Sciences*, 19 (4), 227–233.
- Edwards, M. J., Adams, R. A., Brown, H., Pareés, I. & Friston, K. J. (2012). A Bayesian account of ‘hysteria’. *Brain*, 135 (11), 3495–3512.
- Feldman, H. & Friston, K. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, 4, 215.
- Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11 (2), 127–138.
- (2011). What is optimal about motor control? *Neuron*, 72 (3), 488–498.
- Friston, K. & Frith, C. (2015). A duet for one. *Consciousness and Cognition*, 36, 390–405.
- Friston, K. J., Daunizeau, J. & Kiebel, S. J. (2009). Reinforcement learning or active inference? *PloS One*, 4 (7), e6421.
- Friston, K. J., Daunizeau, J., Kilner, J. & Kiebel, S. J. (2010). Action and behavior: A free-energy formulation. *Biological Cybernetics*, 102 (3), 227–260.
- Friston, K., Mattout, J. & Kilner, J. (2011). Action understanding and active inference. *Biological Cybernetics*, 104 (1-2), 137–160.
- Friston, K. J., Shiner, T., FitzGerald, T., Galea, J. M., Adams, R., Brown, H., Dolan, R. J., Moran, R., Stephan, K. E. & Bestmann, S. (2012a). Dopamine, affordance and active inference. *PLoS Comput Biol*, 8 (1), e1002327.
- Friston, K., Samothrakis, S. & Montague, R. (2012b). Active inference and agency: Optimal control without cost functions. *Biological Cybernetics*, 106 (8-9), 523–541.
- Friston, K., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T. & Dolan, R. J. (2013). The anatomy of choice: Active inference and agency. *Frontiers in Human Neuroscience*, 7.
- Fuchs, T. (2010). The psychopathology of hyperreflexivity. *The Journal of Speculative Philosophy*, 24 (3), 239–255.
- Gallagher, S. (1986). Lived body and environment. *Research in Phenomenology*, 16, 139–170.
- Gallese, V. & Metzinger, T. (2003). Motor ontology: The representational reality of goals, actions and selves. *Philosophical Psychology*, 16 (3), 365–388.
- Hohwy, J. (2007). The sense of self in the phenomenology of agency and perception. *Psyche*, 13 (1), 1–20.
- (2012). Attention and conscious perception in the hypothesis testing brain. *Frontiers in Psychology*, 3.
- (2013). *The predictive mind*. New York: Oxford University Press.
- Hommel, B. (2015). The sense of agency. In P. Haggard & B. Eitam (Eds.) (pp. 307–326). New York: Oxford University Press.
- Hommel, B., Müsseler, J., Aschersleben, G. & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, 24 (05), 910–926.
- James, W. (1890). *The principles of psychology*. New York: Holt.
- Jiang, J., Heller, K. & Egner, T. (2014). Bayesian modeling of flexible cognitive control. *Neuroscience & Biobehavioral Reviews*, 46, 30–43.

- Kilner, J. M., Friston, K. J. & Frith, C. D. (2007). Predictive coding: An account of the mirror neuron system. *Cognitive Processing*, 8 (3), 159–166.
- Kunde, W., Elsner, K. & Kiesel, A. (2007). No anticipation—No action: The role of anticipation in action and perception. *Cognitive Processing*, 8 (2), 71–78.
- Lau, H. C., Rogers, R. D., Haggard, P. & Passingham, R. E. (2004). Attention to intention. *Science*, 303 (5661), 1208–1210.
- Legrand, D. (2011). Oxford handbook of the self. In S. Gallagher (Ed.) (pp. 204–227). New York: Oxford University Press.
- Liang, C. (2015). Self-as-subject and experiential ownership. In T. K. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt am Main: MIND Group. <https://dx.doi.org/10.15502/9783958570030>.
- Limanowski, J. (2014). What can body ownership illusions tell us about minimal phenomenal selfhood? *Frontiers in Human Neuroscience*, 8.
- Limanowski, J. & Blankenburg, F. (2013). Minimal self-models and the free energy principle. *Frontiers in Human Neuroscience*, 7.
- Merleau-Ponty, M. (1945/1962). *Phenomenology of perception* (trans. Colin Smith). London & New York: Routledge & Kegan Paul.
- Metzinger, T. (2004). *Being no one: The self-model theory of subjectivity*. Cambridge, MA: MIT Press.
- (2009). *The ego tunnel: The science of the mind and the myth of the self*. New York: Basic Books.
- (2014). How does the brain encode epistemic reliability? Perceptual presence, phenomenal transparency, and counterfactual richness. *Cognitive Neuroscience*, 5 (2), 122–124.
- (2017). The problem of mental action. Predictive control without sensory sheets. In T. Metzinger & W. Wiese (Eds.) *Philosophy and predictive processing*. Frankfurt am Main: MIND Group.
- Norman, D. A. & Shallice, T. (1986). Consciousness and self-regulation. In R. J. Davidson, G. E. Schwartz & D. Shapiro (Eds.) *Consciousness and self-regulation* (pp. 1–18). New York: Springer.
- Ondobaka, S. & Bekkering, H. (2012). Hierarchy of idea-guided action and perception-guided movement. *Frontiers in Psychology*, 3, 579.
- Pezzulo, G. & Cisek, P. (2016). Navigating the affordance landscape: Feedback control as a process model of behavior and cognition. *Trends in Cognitive Sciences*, 20 (6), 414–424.
- Pezzulo, G., Rigoli, F. & Friston, K. (2015). Active inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology*, 134, 17–35.
- Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology*, 9 (2), 129–154.
- Sass, L. A. & Parnas, J. (2003). Schizophrenia, consciousness, and the self. *Schizophrenia Bulletin*, 29 (3), 427–444.
- Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, 17 (11), 565–573.
- (2015). The cybernetic Bayesian brain: From interoceptive inference to sensorimotor contingencies. In T. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt am Main: MIND Group. <https://dx.doi.org/10.15502/9783958570108>.
- Seth, A. K., Suzuki, K. & Critchley, H. D. (2011). An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology*, 2.
- Stock, A. & Stock, C. (2004). A short history of ideomotor action. *Psychological Research*, 68 (2–3), 176–188.
- Taylor, J. E. T. & Witt, J. K. (2014). Altered attention for stimuli on the hands. *Cognition*, 133 (1), 211–225.
- Voss, M., Ingram, J. N., Haggard, P. & Wolpert, D. M. (2006). Sensorimotor attenuation by central motor command signals in the absence of movement. *Nature Neuroscience*, 9 (1), 26–27.
- Wulf, G., McNevin, N. & Shea, C. H. (2001). The automaticity of complex motor skill learning as a function of attentional focus. *The Quarterly Journal of Experimental Psychology: Section A*, 54 (4), 1143–1154.
- Zahavi, D. (1994). Husserl's phenomenology of the body. *Etudes Phénoménologiques*, 10 (19), 63–84.