Thomas Metzinger

Minimal Phenomenal Experience

The ARAS-model theory: Steps toward a minimal model of conscious experience *as such*

This manuscript provides a short summary of an interrelated set of ideas presented in my third Carnap Lecture, given on 9 March 2018 at the Ruhr-Universität in Bochum, Germany. <mark>The current manuscript is only for purposes of discussion, it will not be submitted to any peer-reviewed journal</mark>; but all comments, relevant phenomenological case-studies, and constructive criticism are welcome. Some parts of it, together with new material, will be submitted in 2019, to a new academic open-access journal titled *Philosophy and the Mind Sciences* edited by Sascha Fink, Wanja Wiese, and Jennifer Windt (see for example Dresler & T. Metzinger, in preparation; Gamma & T. Metzinger, forthcoming; T. Metzinger, 2019, submitted). A simple 80-minute video of the original talk with slides is [electronically available](#). The current manuscript is a work in progress and part of a larger research project.

1. Introduction

In 2009, Olaf Blanke and Thomas Metzinger asked two questions: What are the minimally sufficient conditions for the appearance of a phenomenal self, that is, the fundamental conscious experience of *being someone*? What are necessary conditions for self-consciousness in *any* type of system (Blanke & T. Metzinger, 2009, p. 7)? One of their results was that bodily agency is a causally enabling, but not a constitutive condition, for phenomenal selfhood (Blanke & T. Metzinger, 2009, p. 13). Blanke and Metzinger then introduced the concept of "minimal phenomenal selfhood" (MPS), the central defining features of which are, (i) a globalized form of identification with the body as a whole (as opposed to mere ownership for body parts), (ii) spatiotemporal self-location and (iii) a weak, geometrical first-person perspective (1PP).

This paper develops a related strategy for "minimal phenomenal experience" (MPE; this term was originally introduced in Jennifer M. Windt, 2015). It will take not full-body illusions, but the specific phenomenology of "pure consciousness" in meditation as a new entry point. One assumption is that MPE lacks, (i) MPS or other forms of self-consciousness, (ii) time representation, and (iii) a spatial frame of reference. More generally, here it is not the global phenomenology of identifying with the body as a whole that constitutes the empirical starting point, but all reports in which subjects claim that they have actually had an experience of consciousness *as such*. Is there really something like the simplest form of conscious experience? If yes, what are minimally sufficient conditions for MPE in neurotypical humans, and what are necessary conditions for the experience of consciousness as such in *any* type of system?

There are five reasons that make MPE interesting: It could be the common phenomenological denominator that is always present whenever there is conscious experience at all, however rich or limited; second, a concerted attempt to isolate this hypothetical common denominator could lead to interdisciplinary unification by connecting different domains and experimental approaches in a new way; third, by aiming at a minimal model explanation (cf. Section 2) we could arrive at an entirely new theoretical model of conscious experience itself,

providing us with a fresh perspective and possibly even a new understanding; fourth, the MPE-approach could lead to unification within philosophy of mind itself, by elevating comparative and transcultural philosophy of mind to a new and more systematic level; and finally, reports about "pure consciousness" experiences are a neglected empirical phenomenon worth studying in itself.

If – as many traditional theories claim[1] - a minimal form of phenomenal experience is present in *all* forms of phenomenal content and functionally underlies *all* types of conscious experience, then developing an evidence-based minimal model of MPE would amount to solving the problem of consciousness *as such*. We could use its neural correlate as a marker for consciousness (T. Metzinger, 2000). Let us call this assumption the "fundamentality constraint": If there really is an all-pervading but mostly implicit kind of phenomenal character which can at times be *made* explicit and which underlies *all* other forms of phenomenal experience, then it follows that the minimally sufficient neural correlate of MPE must be functionally integrated with that of all other conscious states. Put differently, we would then expect "computational bridging principles" connecting information processing in the physical correlate of MPE with all other types of conscious experience. On the other hand, MPE would be the only form of phenomenal content that could in principle be functionally autonomous and perhaps sometimes appear in isolation. I do not endorse the fundamentality constraint at this point. It is an open empirical question if MPE exists and if it satisfies this constraint in neurotypical human beings. All I want to do point out at this stage is a) that the idea has rich precursors in Indian philosophy of mind, and b) that, if true, it would open a completely route for empirical research programs on consciousness.

If this background assumption could be demonstrated to be true, then it could lead to theoretical unification both within empirical consciousness research and across disciplines by connecting different domains and experimental approaches in a new way. Second, a minimal model would also provide a novel and deeper *understanding* of consciousness because the specific "representational ideals" governing the construction, analysis, and evaluation of theoretical models under the proposed strategy of minimalist idealization provide us with a new and more parsimonious *perspective* on the target phenomenon (Weisberg, 2012; section 6.2). At the same time, solving the problem of pure awareness would also be of highest relevance for comparative and transcultural philosophy: "Pure consciousness" has played a significant role in non-Western philosophy of mind, in particular, Advaita Vedanta and Buddhist theories of mind (cf. section 4 and 5.2). Traditional philosophical claims about an "empty" or "contentless" form of experience raise a problem for the currently leading kind of theories about consciousness, namely Representationalism: If this conscious state is not about anything, then there is an experience that is not defined by its content; therefore, Representationalism would be false. Here, first steps towards an investigation of minimal phenomenal experience are provided together with a defence of Representationalism. Thus, on a purely metatheoretical level, a more fine-grained conceptual model of MPE would, by integrating existing philosophical theories across cultural contexts, possess great heuristic and innovative potential.

The central empirical hypothesis of this discussion paper is as follows: For neurotypical human beings, the minimal form of conscious experience is constituted by the content of a predictive model serving to control and regulate the global signal of the *Ascending Reticular Arousal System* (ARAS), which in turn determines the brain's general level of activation. Specifically, the ARAS-signal plausibly is not only the strongest internal source of free energy (Friston, 2010), but at the same time it also mediates cortical arousal which has long been known to be the major causally enabling factor for the instantiation of phenomenal states. The ARAS-model theory says that on its most fundamental level, conscious experience is a model of the hidden cause of the ARAS-signal. However, as explained in sections 2 and 8.2, terms like "ARAS", "ARAS-signal", or "ARAS-model" are here used as *placeholders*, with each of them specifying a whole range of potential empirical research programs.

On a conceptual level, the main claim of this paper is that what has sometimes been called "pure" consciousness – i.e., the hypothetical "essence" of phenomenal awareness *as such* – is really a statistical hypothesis, a non-conceptual *representation*, dynamically modelling the hidden cause of the ARAS-signal (see figure 10 on p. 32). It has an intentional object. "Pure" consciousness really is a representational state, but, interestingly, one that instantiates a unique type of phenomenal character which is typically described as "empty" or non-representational. Part of the idea is that carefully relating the traditional problem of pure consciousness to current work in philosophy of mind and computational neuroscience may give us a new entry point, a fresh perspective on the problem of consciousness. Therefore, one goal is to gain a deeper phenomenological understanding of what, over the centuries, has often been reported as "bare awareness" or "conscious*ness* as such" (and also *why* it has often been falsely reported as lacking content). A second goal is to find out whether the philosophical problem of defining "minimal phenomenal experience" (MPE) can be approached by conceptually developing a novel multidimensional construct. We need a new and more fine-grained psychometric instrument targeting the experience of "pure" consciousness – which, in humans, is implemented by what was provisionally called the ARAS-model above.

In addition, the ARAS-model theory is not simply another initiative towards a classical reductive explanation for the specific phenomenology of pure consciousness. The third epistemic goal behind the current proposal is to test a new explanatory strategy for the problem of consciousness (namely, by constructing a "minimal model" as opposed to isolating minimally sufficient neural correlates and developing information-processing theories for specific forms of conscious content; see section 2). A fourth, intimately related, goal is to arrive at a more fine-grained, evidence-based, and conceptually deeper understanding of traditional phenomenological notions like "bare awareness" or "pure consciousness" by re-contextualizing them within current philosophy of cognitive science. Put differently, the speculative background assumption behind the current proposal is that these traditional concepts could actually *refer* to the simplest forms of conscious experience human beings are capable of and that, given the independent theoretical and metatheoretical motivation just sketched, there is a lot to be learned from them. Therefore, the central aim is to simultaneously develop a new approach and a theory of the *simplest* form of conscious

experience possible: "Minimal phenomenal experience" (MPE). This concept, originally introduced by Jennifer Windt, will be explained in Section 4.

Finally, here is the structure. The following section briefly introduces the strategy of explanation by minimal models. Section 3 sets the epistemic goal in more detail, while section 4 develops an initial set of six phenomenological constraints every successful theory of MPE should satisfy and draws attention to three major methodological obstacles in developing a rigorous research program. Section 5 presents new research on the phenomenon of "lucid dreamless sleep" (LDS) as the empirical possibility of an actual *instantiation* of a minimal model of consciousness in the real world. Section 6 then proceeds to sketch a positive model of MPE, and section 7 lists a series of testable empirical predictions. Section 8 concludes by offering preliminary answers to the questions posed in section 3 and returns to a wider, more philosophical perspective. Here, I have kept all historical references and phenomenological case studies to a minimum, but in later versions I hope to add two appendices containing a collection of phenomenological case studies plus a new psychometric instrument for MPE.

2. Scientific explanation by minimal models

In the present context, developing a "minimal model" of conscious experience is something very benign and modest. For example, given the current state of consciousness science, a minimal model will not yet imply elaborate formal procedures like mathematically delimiting an universality class (Batterman & Rice, 2014). Rather, it can be seen as a heuristic strategy of *minimalist idealization*:

- Constructing and investigating a model of conscious experience that includes only the **core causal factors** giving rise to the target phenomenon;
- including only causal factors that make a difference to the **actual** occurrence and the **essential** phenomenal character itself;
- developing an idealized model of **universal and repeatable features** serving to gradually isolate the fundamental, explanatorily relevant, structurally stable properties that underlie disparate forms of conscious experience.

The opposite of a minimal model would be a *maximal* model (i.e., a model which would be "ideal" in a very different sense), a theory which would capture all the behaviour, both repeatable and unrepeatable, of a given system. For complex systems like the human brain, this traditional modelling strategy is an unobtainable goal; many of the fine details will actually get in the way of scientific explanation and detract from a deeper understanding of the research target (Batterman, 2002, p. 22). However, if the phenomenon to be explained is a large-scale regularity that is largely independent of micro-scale details and features of individual elements, it can prove helpful to identify the *minimal conditions* necessary for the explanatory target to occur, later linking them back to the original phenomenon of interest (for three historically important case studies in ancestral heredity, the ideal gas law in physics, and evolutionary biology, see section 2 in Rice, Rohwer, & Ariew, 2018).

If the systems of interest are conscious systems, the more economical strategy of constructing a *minimal* model would amount to the extraction of mathematically stable

structural features, characterizing abstract patterns within phenomenal dynamics independently of potentially unknowable and fine-grained neuroscientific details, thereby yielding what we are most interested in: universal and reliably *repeatable* properties instantiated by all forms of subjective experience. I think that individual episodes minimal phenomenal experience (MPE; for more see section 4) could be instantiations of such an abstract pattern, perhaps even of universal or at least prototypical features. Those states identified by contemplative traditions sometimes called "empty cognizance" or „consciousness as such" may be the most direct route to study such an abstract pattern without any distracting additions.

The strategy of "minimalist idealization" looks for the decisive explanatory and computational factors by first attempting to construct exactly solvable minimal models (Batterman, 2002, p. 37). Therefore, minimal models do not aim at maximal fidelity or completeness; they are not yet mappings to fine-grained functional mechanisms because the accurateness conditions for an actual *representation* of causal pathways are not intended to be very strong (Weisberg, 2012; Chapter 6, section 6.1.2). However, explanation by minimal models may nevertheless involve a phase of *de-idealization* at a later stage, turning them into a continuing research program. I will return to this point at the end. To really solve the problem of consciousness we will ultimately need a theory of MPE – because only a minimal model can give us a deep scientific *understanding* of the essence of phenomenal experience.

3. Setting the epistemic goal: The problem of pure awareness

3.1 MPE: Does pure awareness have a minimal, distinct, and unique phenomenal character *sui generis?*

In the introduction, I said that the concept of minimal phenomenal experience (MPE) refers to "the" simplest form of conscious experience. However, one background assumption guiding the current approach is that the concept of phenomenal consciousness does not have definitional, but probabilistic structure. "Probabilistic structure" means that that a certain conscious experience falls under *MPE* if it satisfies a sufficient number of constraints encoded by *MPE*'s semantic constituents (Margolis & Laurence, 2014). In principle, there could multiple and loosely overlapping sets of sufficient conditions. Ultimately, *MPE* is a cluster concept and "MPE" refers to a phenomenological prototype. This implies that there is no set of necessary and sufficient conditions that allows us to say whether a representational state is conscious or not. Rather, "consciousness" refers to states, regions, and trajectories in a multidimensional space constituted by functional and content-related dimensions – it is a graded and heterogeneous construct, membership comes in degrees (Bayne, Hohwy, & Owen, 2016; for a similar approach to the concept of "mind wandering", see Seli et al., 2018b). The underlying dimensions which give a structure to the space of conscious experience create a complex pattern of family-resemblances, and absolute orderings may not be possible in all cases.

For concrete states, there will be *degrees of prototypicality*, because highly prototypical members have attributes that overlap with most other exemplars of the category, while low-prototypical members have little overlap. Membership in the family of phenomenal states will therefore be graded, and quantitative assessments of prototypicality may actually help us

understand why some phenomenological exemplars better capture the intuitive "essence" of consciousness than others.

However, in searching for concrete examples falling under MPE, one important implication following from this first background assumption is that it may not be possible to isolate *the* simplest form of conscious experience in terms of one unique minimum of dimensionality for the relevant state-space (or an individual state displaying *the* minimal set of characteristic features), but that for a given system multiple low-dimensional states may actually be possible or even co-exist. For example, in a 1.000.000-dimensional state space there might be five different kinds of three-dimensional states, each of them exhibiting a unique phenomenal character.

I have here mostly chosen reports about "pure awareness" or "pure consciousness" from contemplative practitioners as a starting point, but in principle there may be many other relevant kinds of minimal phenomenal experience - for example as caused by subanaesthetic doses of general anaesthetics (Russell, 2013; Sanders, Tononi, Laureys, & Sleigh, 2012) or as experienced by patients moving in and out of coma (Giacino et al., 2002; Steven Laureys, Owen, & Schiff, 2004). My reason for selecting this highly specific subset of contemplative experiences is twofold: If we are interested in arriving at a minimal-model explanation for phenomenal consciousness, it is quite simply the most natural first step to more closely investigate all those states in which subjects actually *report* experiencing the "essence" of consciousness or consciousness *as such*. Second, what makes the phenomenology of certain meditative states interesting from a scientific perspective is that the philosophical phenomenologies typically coming along with them claim a *conjunction* of minimal dimensionality (i.e., an "empty" form of conscious experience lacking internal structure and content) and maximal prototypicality (i.e., an often-unnoticed overlap with most other exemplars of the category of "consciousness").

Here, "often-unnoticed" means that the hypothetical, all-pervasive phenomenology in question is introspectively available in principle, but that we almost never direct our attention there. Hypothetically, there could be two major causes for this situation: First, most of us lack the cultural priors that generate a personal-level expectation that something *can* actually be discovered there; second, because of some specific subpersonal mechanism of "introspective neglect" for the respective region of state-space, a naturally evolved inability to direct our attention there that has not yet been scientifically understood. "Overlap" means that there is an as yet unexplored empirical possibility: The relevant "empty" form of phenomenal experience could be ubiquitous, a constitutive part of *all* other conscious experiences (see Introduction).

From now on, I will use phenomenological notions like "pure awareness" and "pure consciousness" interchangeably. "MPE" refers to the *theoretical* concept of minimal phenomenal experience as it could figure in a minimal model explanation of consciousness. I will here limit myself to one plausibly homogeneous subclass of MPE, to a paradigmatic form of "pure consciousness"-experience as it appears in some forms of contemplative practice. It would have been possible to introduce "MPE_m" in order to flag the fact that I am here using mostly (but not exclusively) reports about certain meditation experiences as a starting point, but I think that, for the purposes of this paper, this move would have complicated matters unnecessarily – by now readers are well aware of the general strategy. The general aim is to develop a conceptually coherent and empirically plausible *minimal* model of conscious experience - a model that takes the phenomenology seriously - and to then see if it maps

onto traditional concepts in contemplative phenomenology. Here are four general questions that any minimal model explanation of conscious experience must answer:

Question 1:

- Is there a form of consciousness that does *not* instantiate minimal phenomenal selfhood (Blanke & T. Metzinger, 2009) or an epistemic agent model (T. Metzinger, 2017; section 2.5)?
    - → Are selfhood and perspectivalness *necessary* conditions for phenomenality?

Question 2:

- Does "pure" consciousness really exist?
    - → Are there any concrete examples of phenomenality *without* intentional content?
    - → Are the conscious experiences which do have content, but which can still count as "pure"?

Question 3:

- Can we develop a minimal model for conscious experience *as such*?
    - → Is there a fundamental and perhaps implicit kind of phenomenal character *sui generis*, which can at times be *made* explicit and which underlies *all* other forms of phenomenal experience?

Question 4:

- Is there one maximally abstract phenomenal property?
    - → Is "pure consciousness a *unique* form of phenomenal character of which we could say that it forms a phenomenological kind of its own?

I will present preliminary answers to these questions in section 6.1.

4. Phenomenological constraints and methodological problems

4.1 Introduction

One major research target is the development of a fully validated psychometric instrument for MPE (cf. Appendix II). I will take three steps. As a first, preparatory step to create such a statistical construct I will now treat MPE as an exclusively *phenomenological* notion. I will attempt to semantically enrich the new concept by extracting six general constraints from both Eastern and Western philosophical traditions, sometimes mixing phenomenological analysis and exegesis, but deliberately leaving aside all of the metaphysical interpretations and theoretical background assumptions frequently offered along with experiential reports by these traditions. Accordingly, we will arrive at a preliminary 6D-model of what has been called "pure" consciousness in the past (readers who are not interested in the details can directly skip to the concluding summary and figures in section 4.6 and 4.7). Second, this six-dimensional proto-concept will be tested against a collection of concrete phenomenological case-studies (Appendix I). Later, this preliminary model of MPE will be developed into a more

fine-grained psychological construct which can eventually be grounded in neuroscientific data and continuously refined (Appendix II).

4.2 The problem of embodied theory contamination

It is important to note that Eastern traditions often do not clearly distinguish between phenomenological, epistemological, and metaphysical readings of concepts such as "pure consciousness". To arrive at a convincing cross-cultural typology of the relevant target states, one would have to separate the level of phenomenological description from the level of interpretation in an unequivocal manner (Wainwright, 1981, 18pp). In addition, the richest and most detailed phenomenological accounts of minimal phenomenal experience clearly come from exactly those spiritual traditions that have over many centuries developed detailed and systematic approaches to the *cultivation* of the conscious states of interest. Theory is therefore intertwined with practice. Thus, a first major methodological obstacle for any research program based on objective third-person evidence lies in the fact that both phenomenological reports and the mental techniques used in the practical induction of the phenomenal states of interest are obviously coloured and conditioned by strong epistemological, metaphysical, and soteriological background assumptions (Melchert, 1977). We can term this first methodological difficulty the problem of "theory contamination".

It is important to understand that theory contamination has strong causal effects on personal as well as subpersonal levels of information-processing and self-representation. For example, if neuroscientific studies targeting MPE were to use monks from the Tibetan Buddhist tradition as subjects, it is not only plausible to assume that their specific belief system will colour their understanding of questionnaires and the taxonomy they employ in describing their own methods and experiential states (Lutz, Slagter, Dunne, & Davidson, 2008; Nash & Newberg, 2013; Fred Travis & Shear, 2010). Conceptually, the Tibetan tradition applies central Buddhist concepts like "dharmakāya" (*chos sku, rdzogs sku*; often translated as "Buddha-nature") and "rigpa" (*rig pa*; "knowledge of the ground"), which in their phenomenological reading may well be directly related to the notion of MPE which is relevant here. But what is more, the cultural context will also be embodied on a subpersonal level because the human self-model functions as a functional link between sociocultural and subpersonal levels of biological self-representation (Frith & Metzinger, 2016, pp. 209–210). Hence deeply engrained background assumptions will also act *internally* as a set of priors and high-level expectations, continuously shaping the causal flow and the way information is processed in the subject's brain - for example, by generating predictions and potentially self-fulfilling prophecies in a top-down manner (in analogy to the "active inference" framework). Let me give a second concrete example for this principle of "embodied theory contamination". It is not only plausible to assume that long-term practitioners of Transcendental Meditation (TM) will produce more phenomenological reports about experiences of "pure consciousness" (J. Bachmann, 2014, 47pp; Severeide, 1990) and "witnessing sleep" than others (see Section 5 on lucid dreamless sleep), but that they will also experience more low-level *bodily* effects like spontaneous episodes of breath suspension (Farrow & Hebert, 1982). Such personal-level and subpersonal-level effects accompanying contemplative practice can be predicted simply because the *teachings* coming along with this specific type of practice explicitly create exactly these expectations plus a strong motivational context, namely, by flagging them as "markers of success". In this first, preliminary step towards developing a proper psychometric

instrument I have tried to deal with this difficulty by extracting phenomenological constraints across a variety of cultural contexts and historical epochs.

4.3 Pure consciousness in contemplative phenomenology

MPE likely is an experience every human being knows. At the same time, most explicit reports come from humankind's contemplative traditions, and I will therefore mostly treat it as a specific and very simple subset of meditative experiences. In his classic monograph *Philosophy and Mysticism* Walter Terence Stace categorizes reports of "pure consciousness" as a specific, namely introvertive, type of mystical experience:

> There would be no mental content whatever but rather a complete emptiness, vacuum, void. One would suppose *a priori* that consciousness would then entirely lapse and one would fall asleep or become unconscious. But the introvertive mystics—thousands of them all over the world—unanimously assert that they have attained to this complete vacuum of particular mental contents, but that what then happens is quite different from a lapse into unconsciousness. On the contrary, what emerges is a state of *pure* consciousness—"pure" in the sense that it is not the consciousness of any empirical content. It has no content except itself. […]
> It is the bare unity of the manifold of consciousness from which the manifold itself has been obliterated. This seems analogous to saying that if from a whole or unity of many parts we could subtract all the parts, the empty whole or unity would be left. (Stace, 1960, p. 86)

Similarly, in a well-known collection of texts, first published by Oxford University Press in 1990 under the title *The Problem of Pure Consciousness*, editor Robert Forman defines "Pure Consciousness Events" (PCEs) as a "wakeful though contentless (non-intentional) consciousness":

> One neither thinks nor perceives any mental or sensory content. Yet, despite this suspension of content, one emerges from such events confident that one had remained awake inside, fully conscious. This experience, which has been called the pure consciousness event, or PCE, has been identified in virtually every tradition. […] The pure consciousness event may be defined as a wakeful but contentless (non-intentional) consciousness. (Forman, 1998)

Interestingly, the same author also endorses something strongly resembling the "minimal model approach" introduced in section 2 above:

> When a biologist seeks to understand a complex phenomenon, one key strategy is to look to at it in its simplest form. Probably the most famous is the humble bacterium *E. coli*. Its simple gene structure has allowed us to understand much of the gene functioning of complex species. Similarly, many biologists have turned to the "memory" of the simple sea slug to understand our own more kaleidoscopic memory. Freud and Durkheim both used totemism, which they construed as the simplest form of religion, to understand the complexities of religious life. The methodological principle is: to understand something complex turn to its simple forms.
> Mystical experiences may represent just such a simple form of human consciousness. […] These PCEs, encounters with consciousness devoid of intentional content, may be just the least complex encounter with awareness *per se* that we students of consciousness seek. The PCE may serve, in short, as the *E. coli* of consciousness studies. (Forman, 1998)

From these classical sources, we can already derive two major phenomenological constraints (PCs):

- WAKEFULNESS (PC1): the phenomenal quality of tonic alertness.

- **LOW COMPLEXITY (PC2)**: Often described as the complete absence of intentional content, in particular the absence of high-level symbolic mental content (i.e., conceptual or propositional thought), but also of sensory content.

If we accept these first two constraints, then a theory of MPE will have to be a theory of contentless, restful alertness – a form of phenomenal experience that in its purest form does not instantiate any intentional properties. From a philosophical perspective, this conjunction of two necessary conditions is simple and straightforward. As a first characterization, it is certainly helpful.

However, please note how even the very first classical quotations presented above introduce a whole range of conceptual ambiguities. If Walter Stace says "It has no content except itself.", then he introduces reflexivity as an additional constraint for a conceptual prototype of pure consciousness; when he speaks of an "empty whole or unity" one is confronted with the problem that on closer inspection the phenomenology of wholeness and the phenomenology of unity are very different (Wiese, 2018). And if Robert Forman defines PCEs by making use of the notion of "(non-intentional) consciousness", this immediately creates the need for a background theory about what "intentional" content really is, how many kinds of mental content there are, and how phenomenal and intentional properties are related. In particular, these conceptual ambiguities leave one specific possibility open: The absence of all empirical content could itself be an appearance, and what subjects (falsely) describe as a "contentless" phenomenal state actually carries an *abstract* form of intentional content. If there is a content that can account for the experience of contentlessness, then Representationalism remains tenable. This is the line I will be pursuing.

From a very different perspective, here is how Jonathan Shear approaches the point:

> Experience of the deepest level, that of pure consciousness awake to its own nature in and by itself, is generally held to be especially important. The defining characteristic of this experience is the *complete absence* of all sounds, tastes, thoughts, feelings, images, and anything else that one can ever imagine. […] But they have in common the idea that it is possible for all empirical content to disappear, while one nevertheless remains awake.

> What then is the experience like? By all accounts it is not *like* anything. One can have it and remember it – one knows that one was not asleep. But one does not remember it *as* anything at all. It is just *itself* – unimaginable and indescribable. The experience itself is extraordinarily abstract. Indeed, it is the logical ultimate of abstraction, since by all accounts it is what remains after everything that can possibly be removed from experience has been removed, while one nevertheless remains awake.

> It should be noted, however, that even if the natural response to this experience is to describe it as "contentless," it is still appropriate to raise the question of whether it is in fact completely contentless. For while experiences […] might be so subtle and abstract that they naturally seem to be completely contentless, they might nevertheless actually have some, albeit very abstract, content. (Shear, 2007, 700; 702-703)

If we are searching for a positive characterization of the target, an additional difficulty is that **LOW COMPLEXITY** already introduces one exclusively negative element. And indeed, if one looks at the relevant literature, it is striking how many phenomenological characterizations of MPE-episodes are of a *negative* kind. (For brevity, I keep things simple here. More examples can be found in the appendices; cf. A-I. ==Again, as this manuscript is only work in progress, all pointers==

). PCEs are described as:

- **Non-sensory**: MPE itself instantiates no perceptual qualities (which also includes "no-thingness", i.e., the absence of the phenomenal property of "objecthood", meaning that there is no subjective experience of distinct *multimodal* objects as integrated from different sensory features and as segmented from a background or perceptual scene);

- **non-motor**: absolute stillness, no motion in space;

- **atemporal**: an absence of temporal experience, no motion in time;

- **non-cognitive**: non-symbolic and non-conceptual, no mind wandering;

- **non-egoic**: no self-location in time, no self-location in space, no quality of agency, no localized unit of identification (i.e., MPE itself is not even characterized by MPS, the minimal phenomenal sense of selfhood);

- **aperspectival**: no epistemic agent model (EAM[2]), no sense of bodily agency, no passive personal-level self-as-subject. Hence the relevant phenomenal character seems not to be a *first-person* character – during its actual occurrence it is not tied to an individual first-person perspective (this feature is often called "non-duality" in Buddhist or Neo-Advaitan literature).

WAKEFULNESS is by far the most frequently mentioned kind of phenomenal character in descriptions of "pure consciousness", and this observation is an important premise for the argument to be developed later. The second most frequent element is an absence of spatiotemporal self-location. "Spatial self-location" means that one experiences oneself as located at a point or a region in a spatial frame of reference (typically as embodied), "temporal self-location" means that one is *present* in a temporal order, in a consciously experienced "now" (Blanke & T. Metzinger, 2009). "Nowness" itself is a form of temporal experience, and this already leads to two conclusions that can guide future neuroscientific research into the underlying correlates: Subjects experiencing the relevant states are not *oriented* in time and space, and the processing of temporal information has been transiently suspended. To demonstrate this second aspect, here are two case studies from Appendix I:

> I experience pure consciousness as a state of unboundedness and total ease and deep relaxation. There are no thoughts, no feelings, or any other sensations like weight or temperature. I just know I am. There is no notion of time or space, but my mind is fully awake and perfectly clear. It is a very simple and natural state. (Alexander, 1988, p. 3)
>
> ... a couple of times per week I experience deep, unbounded silence, during which I am completely aware and awake, but no thoughts are present. There is no awareness of where

---

[2] An EAM is a special layer in the human self-model. It is a conscious internal model of the self as actively selecting targets of knowledge, as a mental agent who stands in epistemic relations (like "perceiving", "believing", "knowing") to the world and to itself (as in "controlling the focus of attention", "reasoning about the world and oneself", or "knowing that one knows"), and as an entity that has the capacity to actively *create* such relations of knowing. The EAM is the origin of our consciously experienced first-person perspective, see T. Metzinger (2017); section 2.5 .

I am, or the passage of time. I feel completely whole and at peace. (Frederick Travis & Pearson, 2000, p. 81)

Against this background of predominantly negative characterizations there is a third conclusion that we can draw about WAKEFULNESS and LOW COMPLEXITY: There will be variable degrees of constraint satisfaction. The concept of "MPE" really refers to a multidimensional space of possible conscious states, a space which may contain different points, regions, or trajectories (Bayne et al., 2016). *Prima facie*, judging from the currently available set of case-studies, there clearly seems to be a *primary* dimension constituted by the phenomenal character we refer to by the folk-psychological concept of "wakefulness", and it is this phenomenological dimension that should become the first target of empirical studies. "Wakefulness as such" seems semantically related to similar phenomenological notions like "mental clarity" (but without mental content), "cognitive lucidity" (but without conscious cognition), or "bare awareness" (without object). What all of these have in common is that they refer to a kind of conscious experience that can clearly come in different strengths and degrees of intensity. "MPE" refers to their common phenomenological denominator.

Let me give two examples. First, it is plausible to assume different degrees of wakefulness phenomenologically characterizing such states (WAKEFULNESS). Subjectively, there should be an introspectively detectable *gradient of tonic alertness*. For example, PC1 could be *higher* during certain meditative or drug-induced states, but lower during episodes of mental fatigue, sleep-deprivation, or when we are close to sleep-onset or fainting.

To make this even clearer, let me briefly introduce a conceptual distinction, which will be explained in detail in section 4.6.1. "Wakefulness" exclusively refers to the *phenomenal experience* of tonic alertness, and not to the functional property of alertness as such or to the concrete physical property of cortical arousal. It is important to never conflate "wakefulness" as referring to a phenomenal property (the conscious experience of sustained, cue-independent, tonic alertness as such) with "wakefulness" as a functional property (for example, as physically realized by the level of cortical arousal; for one possible example of this conflation see figure 7 below). It is logically possible to have functional wakefulness without conscious experience (as likely demonstrated by the existence of a wake/sleep cycle in certain types of coma patients, for example in unresponsive wakefulness syndrome [UWS], tellingly termed *Wachkoma* in German). In addition, it is empirically plausible that the conscious experience of wakefulness can exist in low-arousal states like NREM-sleep (as shown by the existence of lucid dreamless sleep; cf. the references given in section 5 and figures 6 and 11 on p. 25 and 37 below). Hereafter, "wakefulness" therefore only refers to the graded *subjective* experience of tonic alertness.

Second, we can expect different *degrees of absorption* and content superimposition. For example, if the "absence of intentional content" demanded by LOW COMPLEXITY only refers to a complete absence of mind wandering and active high-level thinking, then episodes of clear, effortless mindfulness while still perceiving the physical body and seeing the environment with an entirely quiet mind (as for example in Buddhist walking meditation) will already count as an MPE-experience. "MPE" then simply refers to the region of phenomenal state-space that we get by subtracting from the natural wake state

a. the phenomenology of cognitive agency (as described in T. Metzinger, 2013, 2017),
b. the different types of spontaneous, task-unrelated thought (Fox & Christoff, 2018; Seli et al., 2018a), and
c. all episodes in which perception is completely uncoupled from attention (mere "mind-blanking; cf. Ward & Wegner, 2013).

On the other hand, if "absence of intentional content" also refers to the content of perceptual, sensorimotor, or emotional states, then we make a stronger demand. If there is no sensory or other representational content *whatsoever* superimposed onto a MPE-experience, then one would refer to a much rarer kind of state. A full satisfaction of the LOW COMPLEXITY -constraint means *minimal complexity of reportable representational content* on the level of conscious processing (cf. figure 2 below). Therefore, the prototypical region in state-space will be high on the WAKEFULNESS-dimension, and minimal in terms of content-complexity. I will call this type a "full-absorption episode" from now on.

A full-absorption episode cannot be reported, because the self-referential mechanisms of forming an *auto*biographical memory are suspended. Therefore, only the process of entering into it or of emerging out of it can be faithfully represented in the autobiographical self-model; the episode itself is not a part of the subject's inner life narrative. On the other hand, it may be helpful to distinguish between concurrent ineffability and retrospective report, and we do know from amnesia and trauma research that an absence of reportable conscious experience does not imply an absence of memory. I will come back to this point in section 4.5 below.

Given this background, we can now formulate the second methodological problem: How does one arrive at a conceptually consistent and *positive* phenomenological characterization of MPE? To understand the issue at stake, let us look at a classical example taken from the *Bardo-Thödol*, the "Tibetan Book of the Dead" (8th century CE):

> This brilliant emptiness is the radiant essence of your own awareness. It is beyond substance, beyond characteristics, beyond colour. […] The instant of your own presence is empty, yet it is not a nihilistic emptiness, but unimpeded radiance, brilliant and vibrant. […] Your own awareness, a vast luminous expanse, clarity inseparable from emptiness, is also the Buddha of unchanging light, beyond birth and death. Just to perceive this is enough. If you recognize this brilliant essence of your own awareness as Buddha Nature, then gazing into it is to abide in the state of enlightenment. (Lingpa, 2014, pp. 14–15)

If one takes an intellectually honest, analytical perspective, it is not easy to make sense of passages like these. *Prima facie*, concepts like "brilliant" and "radiant" refer to the sensory phenomenology of brightness, which is a necessary element of all visual experience. "Your own" seems to refer to a phenomenology of ownership, that is, at least a minimal sense of selfhood. "Vibrant" either refers to tactile phenomenology, or – perhaps like the terms "brilliant" and "radiant" – it is used in a purely metaphorical fashion. "Perceiving" is standardly defined as a sensory object-representation unfolding under a consciously experienced first-person perspective. Similarly, if "gazing into" is interpreted as a goal-directed process, it must be an active form of visual perception involving an epistemic agent model (an EAM, as in n. 3 introduced above). The phenomenological configuration described is a highly complex one, it is certainly not "pure" in terms of an absence of intentional content or subject/object-structure. Rather, it seems to involve sensory qualities, an active

phenomenal self, and a first-person perspective. It would be easy to present dozens of very similar descriptions referring to the phenomenology of self-luminous and "pure" awareness, mostly from the Advaita and Tibetan Buddhist traditions (Namgyal, 2006; Rgyal-ba-g'yung-drung, Gurung, & Brown, 2017; K. Rinpoche & Namgyal, 2011). Are such descriptions intelligible?

More charitable interpretations are certainly available, and I will point to some possibilities in the final section. Nevertheless, a staunch analytical phenomenologist might claim that historical sources of this type should be discarded simply because they do not meet today's standards of intelligibility and conceptual clarity. On the other hand, such a conservative attitude smacks of the specific kind of intellectual dishonesty which sometimes goes along with ignoring relevant data for covert strategic or ideological reasons – feigning ignorance by pretending to be analytically rigorous. Very obviously, we have a large number of reports referring to the relevant phenomenological prototype across many centuries. We also have deep philosophical traditions for which such experiences are central, and there are not only great mystics but also literally millions of secular practitioners of meditation in the West who can make at least intuitive sense of phenomenological descriptions like these. Admittedly, such reports are often conceptually problematic, cryptic or inconsistent. Nevertheless, to deliberately ignore humankind's contemplative traditions in the search for a minimal model of consciousness would simply be unphilosophical – if we are interested in the problem of MPE, then they provide by far the best source of fine-grained phenomenological descriptions. Rather, the task is to come up with a *better* positive characterization of MPE.

Hence, before we proceed to extract three further phenomenological constraints from Western philosophy and from a very different historical epoch, we can already state the third classical constraint derived from Eastern contemplative traditions:

- SELF-LUMINOSITY (PC3): a phenomenal property of MPE typically described as "radiance", "brilliance", or the "clear light" of primordial awareness.

SELF-LUMINOSITY is an example of a phenomenological constraint that is frequently mentioned in a specific subset of the relevant literature (e.g., in Tibetan Buddhism), but, as opposed to WAKEFULNESS and LOW COMPLEXITY, is not frequently found in actual phenomenological reports. It may therefore not be a necessary element of the common phenomenological denominator we are looking for and not figure in the final theoretical concept of minimal phenomenal experience. In particular, it could be a metaphorical way of referring to WAKEFULNESS:

> Some people compare this cognizance to a radiant 'thing' that shines with a light like a 'clear light'. It means a sense of being wakeful, a quality of being vividly wide awake, which is empty of any identity and naturally alert (T. Rinpoche, 2000, KP 2598-2600).

There seems to be a third aspect of the relevant phenomenal character that can be described as a concrete, quasi-sensory phenomenology of self-luminosity (sometimes translated as *Eigenstrahlkraft* in German). How can we make sense of it? First, I would propose that the relevant phenomenology of concreteness consists in the fact that the subjective experience of tonic alertness (see section 4.6.3), although clearly mode-neutral and therefore not associated with any interoceptive or exteroceptive sensory modality, nevertheless clearly varies along a dimension of intensity. Something always radiates, but to different degrees.

Second, not only does it come in variable strengths and degrees, it is functionally autonomous and permeates all of phenomenal space. "Self-luminosity" may actually be a visual analogy for the experienced *functional autonomy* of tonic alertness. It may be an attempt to point not so much to a distinct form of phenomenal character or representational content, but to a functional dimension structuring the space of conscious experience, which in turn allows for a relative ordering of global states (Bayne et al., 2016, p. 410). It is also conceivable that an organism may possess an inner representation of this dimension. Self-luminosity may therefore be a way of talking about Wakefulness: Wakefulness is always there whenever there is conscious experience at all, autonomous and largely independent of the every-changing kaleidoscope of surface phenomenology, but it comes in different degrees of intensity and somehow it "radiates."

"Self-luminosity" may therefore be best treated as a visual analogy, as an attempt to express what it is *like*: A consciously experienced visual quality like "redness" cannot be analysed merely as hue plus saturation as it also always varies along a third subjective dimension of "brightness". Phenomenally experienced wakefulness might be a like a domain-general form of brightness which comes in varying degrees of intensity. Zero brightness means zero phenomenal experience. The subjectively experienced brightness of a red object represents sheer stimulus-strength for the organism, that is, the *luminance* of a visually given physical object, and perhaps MPE literally turns out to be a mode-neutral form of brightness. For example, the internal "luminance" in question it could be a dynamical property of some part of the neural body. Therefore, if we want to understand the phenomenal concreteness of mode-neutral and contentless wakefulness, then we have to ask: What, for bare wakefulness, is the non-sensory equivalent of the raw stimulus-intensity caused by the physical property of luminance in a given visual target and which, in the domain of visual consciousness, leads to the experiential quality called brightness? Is there something like "mode-neutral brightness"? What is the stimulus source, its hidden cause? If we read "self-luminosity" as a visual analogy, then the question becomes What exactly is "luminant" here, is it perhaps a part of the neural body? I will give a new answer to this question in section 6.1.

4.4 Consciousness *as such* in Western philosophy of mind

Let us now extract three further semantic constraints, but this time from Western analytical philosophy. Below are two passages from a now-classical paper by G.E. Moore, titled *The Refutation of Idealism*:

> [...] the fact that when we refer to introspection and try to discover what the sensation of blue is, it is very easy to suppose that we have before us only a single term. The term 'blue' is easy enough to distinguish, but the other element which I have called 'consciousness' – that which a sensation of blue has in common with a sensation of green – is extremely difficult to fix. [...] And in general, that which makes the sensation of blue a mental fact seems to escape us; it seems, if I may use a metaphor, to be transparent – we look through it and see nothing but the blue; we may be convinced that there *is something*, but *what* it is no philosopher, I think, has yet clearly recognised. (Moore, 1903, p. 446)

> [...] the moment we try to fix our attention upon consciousness and to see what, distinctly, it is, it seems to vanish: it seems as if we had before us a mere emptiness. When we try to introspect the sensation of blue, all we can see is the blue: the other element is as if it were diaphanous. Yet it can be distinguished if we look attentively enough, and if we know that there is something to look for. (Moore, 1903, p. 450)

First, it is noteworthy that Moore actually uses the term "emptiness", which plays a great role in Buddhist philosophy (cf. the preceding quotation). Second, it is noteworthy that he also uses a visual metaphor to describe the mode-neutral phenomenal character in question. His paper is the *locus classicus* for the concept of "phenomenal transparency", a property of some conscious states which he called "transparency" or "diaphanousness". Today, a standard definition of "phenomenal transparency" is that it essentially consists in only the *content properties* of a conscious mental representation being available for introspection, but not the fact that it also has non-intentional or "vehicle-properties". In other words, it is not experienced *as a representation*. Introspectively, we can access its content, but not the content-formation process itself. The carrier is invisible. Often, it is assumed that transparency in this sense is a property of *all* phenomenal states. But of course, the standard assumption is incomplete – *opaque* phenomenal representations also exist (whereas unconscious states are neither transparent nor opaque in this sense). The most notable phenomenological examples of opaque state-classes are consciously experienced thoughts: We experience them as mind-dependent and internally constructed, as mental representations that could be true or false. Apart from that, some types of emotions, pseudo-hallucinations, and lucid dreams are also subjectively experienced *as representational processes*. They make the possibility of *mis*representation introspectively available. There is therefore a spectrum between phenomenal opacity and phenomenal transparency, and any given content can vary along this spectrum. Phenomenally opaque processes sometimes appear to us as deliberately initiated cognitive or representational processes. However, they can also appear as automatic or spontaneously occurring, as limited or even global phenomenal simulations, they can draw us into a manifest daydream, and they frequently seem not to be under the experiential subject's control (for more, see T. Metzinger, 2003a, 2003b, 2014).

Moore presents a phenomenological argument against the transparency of consciousness *as such*. Many authors have read Moore as if he were talking about the "transparency of qualia"[3]. However, the relevant point is not about the transparency of awareness or qualia, but rather the fact that consciousness as such can be *made* phenomenally opaque (for a well-researched and substantial discussion, see Hellie, 2007). Moore's much deeper point is that conscious experience *as such* has a unique phenomenal character *sui generis* (namely, a second-order "awareness-of", a relational "signature of knowing"), and that this character can sometimes be detected by introspective attention. In fact, Moore's self-stated goal in introducing "transparency" is actually "to try to make the reader *see* it" (Moore, 1903, p. 450). His important phenomenological claims are therefore:

- **(M1)** There is one most general phenomenal property, which is shared by all sensory qualities:
  → "Consciousness" *as such*, pure awareness, phenomenality *per se.*
- **(M2)** Under standard conditions, this global property is "transparent" (Moore, 1903, p. 446), a "mere emptiness" or "diaphanous":

---

[3] Cf. Block (1996, p. 26-27), Kind (2003, p. 229), Tye (1992, p. 160), Tye (2002, p. 139), Kennedy (2009, p. 574-577), Speaks (2009, p. 539), Stoljar (2004, p. 341). For an excellent, well-researched and substantial discussion containing these and further references, see Hellie (2007).

- → It is not explicitly experienced, but it *can* be.
- **(M3)** This property is *evasive*:
  - → It "seems to vanish" under attentional agency; i.e., if we actively try to "fix our attention" on it (Moore, 1903, p. 450), thereby creating an EAM.
- **(M4)** This property can *become phenomenally opaque*, under two conditions:
  - → we look "attentively enough",
  - → and we "know" that there is a possible object for introspective attention (Moore, 1903, p. 450).
- **(M5)** It is difficult to simultaneously direct and sustain introspective attention to the global property in question and to concrete perceptual qualities:
  - → Moore found no philosopher in the literature who was "able to hold *it* and *blue* before their minds and to compare them, in the same way in which they can compare *blue* and *green*" (Moore, 1903, p. 450).
- **(M6)** Awareness is relational and consciousness is a second-order epistemic process:
  - → Having a sensation is an awareness *of* something, and consciousness is the knowledge that this awareness currently exists. In his own words:

> The true analysis of a sensation or idea is as follows. The element that is common to them all, and which I have called 'consciousness,' really is consciousness. A sensation is, in reality, a case of ' knowing ' or 'being aware of' or 'experiencing' something. When we know that the sensation of blue exists, the fact we know is that there exists an awareness of blue (Moore, 1903, p. 449).

- **(M7)** Consciousness is meta-awareness:
  - → The *type* of epistemic relation is identical – "awareness of" is the same relation in sensation as it is in becoming aware of this awareness. Once again, in Moore's own words:

> This relation is just that which we mean in every case by 'knowing '. To have in your mind ' knowledge' of blue, is *not* to have in your mind a ' thing ' or 'image ' of which blue is the content. To be aware of the sensation of blue is not to be aware of a mental image- of a ' thing,' of which ' blue ' and some other element are constituent parts in the same sense in which blue and glass are constituents of a blue bead. It is to be aware of an awareness of blue; awareness being used, in both cases, in exactly the same sense (Moore, 1903, p. 449).

One could interpret Moore as saying that a) an aperspectival and non-agentive form of meta-awareness is co-instantiated with all forms of conscious perceptual knowledge; that b) the second-order epistemic relation of an awareness-of the current existence of an awareness-of some specific perceptual phenomenal character is mostly unnoticed (i.e., transparent), but c) that we can make it phenomenally opaque by attending to it in the right way. In addition, we can extract three further phenomenological constraints to semantically enrich our six-dimensional working concept of MPE:

- INTROSPECTIVE AVAILABILITY (PC4): we can sometimes actively direct introspective attention to consciousness *as such* and we can distinguish possible states by the degree of actually ongoing access.

- EPISTEMICITY (PC5): The phenomenal experience of knowing, which comes in degrees and can also be described as the subjective quality of *confidence*. To avoid any misunderstandings at this point: The *phenomenology* of knowing can exist without there being knowledge, PC5 is an exclusively phenomenological constraint, and it can be satisfied by intuitions, hallucinations, illusions of control, complex delusional states, and all other forms of consciously experienced misrepresentation.
- TRANSPARENCY/OPACITY (PC6): like all other phenomenal representations, MPE can vary along a spectrum of opacity and transparency.

Let us take a closer look at these constraints but using concrete examples. INTROSPECTIVE AVAILABILITY has been endorsed not only by G. E. Moore, but by many Eastern contemplative traditions (cf. AI-@@). However, it is important to always clearly distinguish between the dispositional property of *availability* for introspective attention and *actually ongoing* introspective access to the relevant phenomenal character. "Availability for introspective attention" is a functional property a given internal state may or may not have, whereas ongoing introspective access presupposes the capacity for active attentional control, at least of the beginning of an episode. This actually ongoing access may then change the overall state.[4] At the same time, it is also clear that the phenomenology of ongoing introspective access is only a transitory phase, and that it is not itself an element of the relevant conscious state, which is devoid of subject/object-structure and any form of conceptual or propositional mental content. Here is how the relevant transition is described from a Tibetan Buddhist perspective:

> The moment you recognize, it is already seen. There is nothing extra remaining that you missed. This is not like space looking at itself, because space does not see anything. When your mind, which is cognizant, recognizes itself, you immediately see that there is no 'thing' to see. It is already seen in the same moment. At that very moment there is no thought, because the present thought has naturally vanished (T. Rinpoche, 2013, p. 79).

> In this description of contemplative phenomenology, WAKEFULNESS and EPISTEMICITY are interpreted as aspects of a self-sustaining, functionally autonomous process that precedes introspective access: Empty cognizance is an original state that is never lost. It merely needs to be known. We need the knowledge of self-existing empty cognizance — the *knowing itself* to be empty and cognizant (T. Rinpoche, 2000, KP 1796-1797).

There seem to be three elements here: First, a functionally autonomous state of subjective confidence, of "knowing without an object"; second, an episode of introspective attention directed at this state; and third the emergence of a new phenomenal state which *integrates* both elements. If MPE were to be the first element - in terms of minimally complex, full-absorption states characterized by a full satisfaction of the LOW COMPLEXITY -constraint (as discussed above) - , then it would be a functionally autonomous kind of internal process that could in principle be actively introspected, but mostly isn't. Let us assume the process in question would be the process by which the central nervous system activates itself, creating the property of tonic alertness. If attended to, the original process could be "made reflexive"

---

[4] The concept of "introspective attention" is only indirectly related to classical theories about "metacognition" Flavell (1979); Metcalfe and Shimamura (1996) Introspective attention is non-conceptual, a subsymbolic process that has nothing to do with "beliefs", "thinking about thinking", or declarative knowledge about one's own states. Rather, it is related to the optimization of precision expectations relative to internal stimuli, cf. Wiese and Metzinger (2017).

or "ignited", thereby causing a new state to appear – yielding a new type phenomenal character that we might provisionally call "self-aware wakefulness".

INTROSPECTIVE AVAILABILITY can therefore be interpreted as a second *functional* dimension (as opposed to a content-related one). Nevertheless, it has to do with what Bayne, Hohwy and Owen (Bayne et al., 2016, p. 409) call a "gating" of the contents of consciousness: MPE itself may be a very simple, unstructured type of phenomenal content, but there is variability along a dimension of actually ongoing introspective access and it might open the organism to other kinds of conscious contents, for example by making it sensitive to incoming perceptual stimuli. or enabling orientation in time and space. Adding introspective attention would transiently create a sense of effort, a phenomenal self, and an internal, self-directed first-person perspective. This could then be followed by brief periods of effortless meta-awareness (i.e., a reflexive form of "empty cognizance") in which the first-person perspective briefly dissolves until the next train of apparently spontaneous, task-unrelated thought ensues. The emerging picture is in full accordance with recent models of what actually happens during the process of mindfulness meditation (Hasenkamp, 2018; Hasenkamp, Wilson-Mendenhall, Duncan, & Barsalou, 2012).

EPISTEMICITY can now be read as a prediction about the phenomenal character of more complex episodes in which introspective attention has been transiently added to a simple state of MPE. Attentional agency certainly creates the phenomenology of a knowing self, an EAM (T. Metzinger, 2017), but such states would then still be characterized by an explicit subject-object structure. Here, the additional idea is that there may be a *non-agentive* variety of "knowingness" in low-complexity MPE-states lacking this structural feature, characterized by a more globalized and effortless form of meta-awareness. MPE is not a mystical experience, but in research on mystical experience, this phenomenological feature has often be described as a unitary state exhibiting a "noetic" phenomenal quality, while still lacking any specific form of content (Wainwright, 1981). In such states one would predict to find a phenomenal signature of knowing which a) has no object (i.e., a form of knowing or subjective confidence without content), and which b) has been globalized (i.e., which is not tied to a localized model of a knowing self any more). Here is one phenomenological description from an advanced meditator:

> [...] "there's no personal point of view, it's the world point of view, it's like the world looking, not ME looking, the world is looking." (quoted from Ataria, Dor-Ziderman, & Berkovich-Ohana, 2015, p. 141)

Clearly, MPE is neither an internal model of a subject, nor a model of an object. Philosophically, this creates a conceptual tension: If we want to follow Moore in interpreting the concept of "awareness" as an epistemic relation, then it becomes difficult to specify the subject-argument place. What exactly are the relata of this special relation? What is the known, and what is the knower? It may turn out that the experience itself isn't really contentless, but only described as such, and it may also be non-egoic in the sense that no knowing self (i.e., an EAM) is part of its content. But, under an epistemological analysis, who or what is the epistemic subject? For now, it could only be the biological organism as a whole.

If Moore is right, then MPE actually *is* an epistemic relation ("awareness-of"). If MPE is successfully introspected (i.e., without – as Moore tells us - trying to "fix" attention on it in an

effortful, agentive way), then it is natural to expect a distinct phenomenal character of knowing. However, in this very special case there would be no correlated epistemic object and all the other negative constraints listed in section 4.2 would be satisfied as well. "EPISTEMICITY" therefore refers to the mere phenomenal signature of knowing, i.e., an experience of subjective *confidence* or certitude. MPE could then be the special case presented as a non-directed, non-agentive and aperspectival "knowingness" with or without perceptual object, depending on the degree of absorption discussed above.

This is interesting, because there seems to be a graded, but distinct phenomenal character of "certitude" or subjectively experienced confidence, which we can analyze as a non-conceptual representation of the "degree of belief", "plausibility", "trustworthiness", or "reliability" assigned to an inner or outer perception, or any other consciously experienced epistemic state. We could also term it the "phenomenology of validity", and it clearly is an essential aspect of MPE. It can be found in complex hallucinations, various psychiatric syndromes, or during epileptic seizures, as a result of cortical dysfunction (Picard, 2013). What makes the phenomenal character of subjective confidence especially interesting from a philosophical point of view is a) it directly relates to centuries of theorizing about consciousness as "knowing that one knows", and b) that there are now unifying conceptual frameworks describing confidence as *Bayesian probability*, bridging different fields of psychology and neuroscience and unifying disparate research perspectives (Meyniel, Sigman, & Mainen, 2015). EPISTEMICITY could be the way Bayesian probabilities and model evidence are represented internally, by the system itself and on the level of conscious processing. MPE is a highly relevant target for research, because a better understanding of EPISTEMICITY may help us to study this aspect in isolation.

However, an additional phenomenological aspect of PC5, the epistemicity-constraint, is that MPE-states are sometimes described as reflexive, and sometimes not. That is, while awareness as such has no external intentional object it is sometimes described as having *itself* as an object, of being self-reflexive in a way that is non-conceptual and which does not necessarily involve the experience of a substantial, separate self (MacKenzie, 2008). If there was such an intrinsic, and fundamental form of reflexivity, then one could imagine it either as co-existing with specific content or as a pure stand-alone phenomenon. The issue of a fundamental, non-egoic form of reflexivity built into consciousness itself has been discussed for centuries by Buddhist philosophers (Finnigan, 2018; Williams, 1998) and, following Brentano's classical idea of a single state simultaneously directed at an intentional object and itself (Brentano, 1973, 153p, 2012, 25p), many fine-grained conceptual models have been developed in more recent analytical philosophy of mind (Peters, 2013). Currently, the issue is increasingly gaining attention (Kriegel 2019, Thompson 2011, Ganeri 2018, Zosipovic 2019). Please note how I am here only interested in the phenomenal character of intrinsic reflexivity itself, in the reported experience of an explicit, introspectively available, non-egoic, non-conceptual, and exclusively self-directed awareness. This character is described in the following case-study:

> It's absolutely alive to itself, fully conscious, and it's conscious of itself as No-thing, colorless, featureless, totally empty and totally full. The Space is conscious of itself at the near end, *but it has no personal characteristics* (Harding, 2000, p. 20).[5]

Finally, TRANSPARENCY/OPACITY shows why MPE is such an interesting and limiting case for theories of consciousness. Please recall that a full satisfaction of LOW COMPLEXITY demanded a complete absence of intentional content for MPE, and that "transparency" was defined on the representational level of description. On this level of analysis, transparency is a property of phenomenal representations, and is brought about by the fact that only their intentional or content-properties are available for introspection, but not properties of the physical carrier itself (i.e., of what often is called the "vehicle" in traditional philosophical terminology). From this, it follows that for the special case of a putatively contentless state like MPE, if it really were transparent, there would be *nothing* introspective attention could ever discover – there would simply be no intentional properties that could ever be available for introspection. If, however, MPE were phenomenally opaque (if it was experienced *as* a representation), then

---

[5] What makes EPISTEMICITY even more interesting – and potentially relates it to TRANSPARENCY/OPACITY – is the additional phenomenological fact that, while low-complexity MPE-states are often described either as states of "pure knowing" (cf. A-I 2.1.1), apparently very similar configurations are sometimes also later reported as conscious episodes of "pure being" (cf. A-I 2.1.2).

What is the relationship between "pure knowing" and "pure being", if once again we strip away all epistemological and metaphysical interpretations of the relevant reports, staying with just the phenomenal character itself? There obviously exists a globalized, non-agentive phenomenology of epistemicity *as well* as a phenomenological counterpart of pure onticity. The first type of phenomenal character (i.e., globalized epistemicity without an object, often described as "empty cognizance", can already be found in first the quotation presented above. Let us look at two examples reporting the experience of a globalized form of "empty onticity", of mere existence *as such*, with the first one coming from William James, The Father of American psychology himself:

> During the syncope, there is absolute psychic annihilation, the absence of all consciousness; then at the beginning of coming to, one has at a certain moment a vague, limitless, infinite feeling – a sense of *existence in general* without the least trace of a distinction between the me and the not-me. James (1950, p. 273)

Here is another report of regaining of the process of regaining consciousness following a severe accident, as an illustration what I mean by the "phenomenology of onticity", again without any form of subject/object-dichotomy:

> There was something, and the *something* was not the nothing. The nearest label for the *something* might possibly be "awareness," but that could be misleading, since any awareness I'd ever had before the accident was *my* awareness, my awareness of one thing or another.
> In contrast, this *something*, if it be called awareness, had no *I* as its *subject* and no content as its *object*. It just was. Sullivan (1995, p. 53)

This is puzzling and interesting at the same time. There seem to be borderline-phenomenologies connecting the region of "pure awareness" with that of "pure being" in phenomenal state-space, linking the experience of knowing *as such* with the experience of being *as such*. One logical possibility is that "epistemicity" as well as "onticity" could refer to almost the same kind of global state. This would be again be characterized by the absence of discursive thought, perceptual content, spatiotemporal self-location, agency etc. (as already listed above, see AI-I-@@ for further examples). The only difference would be that in the first category this global state is still experienced as *representational* (i.e., as a contentless form of *awareness-of*) while in the second case it isn't. Perhaps, the first type of state is reflexive (and therefore experienced as self-*representational*), while the second one isn't. In other words, the relevant difference would then be the difference between a phenomenally opaque and a phenomenally transparent representation of the inner process creating tonic alertness.

we are forced to conclude that PC1 (WAKEFULNESS) and PC3 (SELF-LUMINOSITY) actually refer to an introspective experience of certain properties of the physical carrier itself. For example, such non-intentional "vehicle properties" could simply be earlier processing stages in the brain, certain dynamic aspects of the very construction process that creates the level of phenomenal representation in the first place. As we have many reports of MPE-like states across the centuries and across cultures, it is now plausible to conclude that at least some of these reports may actually be reports about an entirely *non-intentional* process, not merely a mode-neutral one.

Now imagine that you would make MPE phenomenally opaque by just the right kind of "looking attentively enough" while at the same time not trying to agentively "fix" attention on it, exactly in the way Moore described. A third conclusion is that if the fundamentality-constraint is satisfied and MPE really underlies all other forms of conscious content, then – if it were made opaque – we would expect a global phenomenal effect. In richly superimposed states involving perception of body and environment everything should then begin to be experienced as representational, as a virtual model lacking an inherent, mind-independent existence of its own, not as a world but as the content of an epistemic space. In LOW COMPLEXITY-situations this would lead to the phenomenology of "empty cognizance" discussed above. If the internal model of our epistemic space was transparent, then its content would simply appear as "real" and directly given, as the world we live in. In LOW COMPLEXITY situations it would create an experience of "pure being" (see note 7). WAKEFULNESS is the phenomenally opaque counter piece to the transparent, conscious experience of realness. In full-absorption episodes there would be neither introspection nor reportability, just a first-order model of the non-intentional, physical carrier as such. In absorbed, but weakly superimposed states already permitting introspective access the phenomenological prediction is that phenomenal quality of EPISTEMICITY (PC5) would be co-instantiated with WAKEFULNESS (PC1) and SELF-LUMINOSITY (PC3) – an experience of subjective confidence related to knowing wakefulness and the "mode-neutral brightness" discussed above.

This leaves us a with very first, semantically enriched 6-D concept of "minimal phenomenal experience". MPE is a representation of wakefulness, but it comes in variable degrees of constraint satisfaction. It can be coupled with introspective access or not, it can be reflexive or not, and it may be characterized by a variable degree of transparency. Obviously, it can also *co-exist* with many other forms of conscious contents and is mostly unnoticed. In a prototypical form, it would be characterized by a mode-neutral form of bare and functionally autonomous "self-luminous wakefulness". The phenomenological prototype would be characterized by a complete absence of time-representation as well any form of sensory, motor, and high-level cognitive content. Further, it would have an absence of low-level embodiment in the sense of spatiotemporal self-location and affective background, as well as of higher levels of selfhood like attentional control and cognitive agency. MPE *as such* is atemporal, selfless, and not tied to an individual first-person perspective.[6]

---

[6] If all of the above is correct, MPE is a counter-example to the "projective geometry"-model of consciousness formulated by Williford, Bennequin, Friston, and Rudrauf (2018) Because of its fundamentally aperspectival nature, it is a phenomenal state not covered by this theory. If one additionally assumes that MPE actually is a

4.5 The problem of performative self-contradiction

This point leads to the third and, in my view, most difficult methodological problem, which is at heart an epistemological and logical problem. MPE-reports often constitute a "performative self-contradiction". A performative self-contradiction arises whenever the propositional content of a statement contradicts the presuppositions of asserting it (Apel, 1998). Recall that the first problem was the potentially high degree of "theory contamination" in phenomenological reports, the second difficulty was to arrive at a genuinely *positive* characterization of the experience itself. All three problems are interrelated, as we can now see. Because it cannot be introspectively validated from the subject's individual first-person perspective, the "purity" or "minimality" or the alleged "non-duality" of MPE itself is a *phenomenally indeterminate* aspect of conscious experience.[7] Hence, because they involve a performative self-contradiction, autophenomenological reports of MPE are highly problematic as data for a statistical evaluation in scientific research programs.

As an example, consider evaluating an idealized, hypothetical autophenomenological report like this:

> "I participated in a silent, 9-day, Buddhist-style meditation retreat. During the second morning meditation of the fifth day, I experienced pure awareness."

This gives rise to the following questions:

- If *you* weren't there, why do you have an autobiographical memory of the episode?
- If it was *timeless*, why do you know when it began or how long it lasted?
- If there was no *self-location in space*, why do you know where it happened?

More fundamentally, the question is: How could a selfless episode of consciousness ever turn into an element of the autobiographical self-model, into a part of your *own* inner life-narrative? I have sometimes called this the "Refrigerator Light Problem": You know that the light inside the refrigerator goes off when you close the door. But whenever you try to validate this knowledge and very carefully open the refrigerator – as it were, secretly peeping into it – you immediately and automatically turn on the light. Technically speaking, introspective knowledge of one's own conscious states presupposes the activation of an EAM (e.g., a mental model of an epistemic agent actively controlling the focus of introspective attention, see note 7) and a minimal sense of selfhood (for example as being located in a temporal frame of reference, of something being experienced in the present moment, plus a temporally extended introspective process that leads to the new phenomenal state). How could one have an episodic memory of an interval where these two presuppositions were not

---

dynamic, Bayesian model of an internal bodily state, then it can be described as bodily self-representation from a God's eye point-of-view.

[7] One philosophically interesting point (which will not be elaborated here) is that MPE is characterized by a deep form of phenomenal indeterminacy, and that neurotypical human beings are introspectively blind to this important fact concerning the foundation of their own conscious experience. It is phenomenally indeterminate in that introspectively, it is neither the case that MPE *is* the simplest and most basic form of phenomenal character we know, nor is that *not* the case. Yet most of us seem to be blind to this neither-nor-ness. "Indeterminacy blindness for MPE" means that human beings are usually unaware of the fact that they are introspectively blind to truly *fundamental* phenomenal facts regarding the indeterminate *foundation* of their very own phenomenal experience as such.

met and still instantiate the phenomenal property of "ownership" relative to this episode? In the words of Rocco Gennaro:

> For one thing, the practitioner is clearly taking the mental state to be *her own* as opposed to someone else's. For another thing, it is difficult to understand how practitioners can later remember and describe these events without having employed conscious I-thoughts during the alleged PCE, that is, without having experienced the event as one's own [...]. Moreover, episodic memory is a form of declarative memory, described as a "remembering *that*" something happened. Surely such memory involves the application of the concepts contained in the "that-clause." (Gennaro, 2008; sections 3.1 and 3.2)

The problem is not only one of logical syntax, created by the language in which such reports are given, and which is obviously structured through a depth grammar always involving subject and predicate. It is the process of episodic memory retrieval itself that turns something non-conceptual into something conceptual and which superimposes a subject-object-structure onto something that was originally a non-dual state. In addition, there are many ways in which subjective reports of meditation experience could also involve artefacts created by what I have called "embodied theory contamination" above and by what elsewhere has been termed "introspective sensitization", systematically transforming the original target into a more complex altered state of consciousness (Pekala, 1991, p. 61).

This leads to three important conclusions. First, prototypical MPE is strictly ineffable. *As such,* it can never become the content of autobiographical memory, and it cannot therefore be reported. At this point, however, it is of utmost importance to avoid a frequent misunderstanding: To say that MPE is "ineffable" does not imply that the relevant phenomenal character is in any way dramatic, something akin to a mystical experience, or something experienced as rare and highly salient. Quite the contrary: So far, everything points to the possibility that MPE is something extremely subtle and generic, perhaps abstract in a non-conceptual sense - not a "peak experience", but rather a "valley experience."

Second, all reports about MPE are necessarily reports of more complex configurations; they are never reports about a truly minimal form of conscious experience occurring in isolation. They do not refer to full-absorption episodes, but to states in the vicinity of full absorption or to the phenomenology of transition. Third, and relevant to the empirical side of this research project, there may be different memory systems at play in the generation of the relevant reports. For example, it is conceivable that certain parts of the brain record selfless episodes while others later re-present them in an autobiographical format, as if they had been experienced by a self. We can do justice to these conclusions by formulating the "Superimposition Principle" (SP):

- **SP**: Reportable episodes of minimal phenomenal experience always refer to phenomenal contents that has been *superimposed* on MPE, the apparently contentless experience of wakefulness as such.

We do not yet know what the causal mechanism enabling this superimposition in the brain is. Nor do we know what the abovementioned hypothetical "neurocomputational bridge" connecting the correlates of MPE with specific contents in the global workspace of consciousness really is or if it exists at all. Nevertheless, the key observation is that there will

be levels of complexity for the superimposed content as well as degrees of depth for the actual superimposition itself. Accordingly, scientific research programs targeting MPE will have to begin by looking for low-complexity situations characterized by shallow content superimposition. Sometimes the background of MPE may even become so dominant that it turns into the foreground of experience. We can imagine situations where the previous foreground has been minimized to the degree that there is no explicit experience of spatial or affective embodiment, no conceptual or propositional thought whatsoever, and only a minimal interoceptive background plus a minimal "self in the act of attending" (see section 5.2 and figure 2 below). Perhaps such "de-immersed" low-complexity situations could sometimes lead to a transition into the most general Gestalt switch conceivable in the phenomenal field, with MPE turning into the (contentless) foreground, creating an explicit experience of emptiness that envelops whatever residual form of self-representation still exists, but without thereby *completely* eliminating it. Conceivably, configurations of this type could then form the functional basis for the construction of an autobiographical self-model.

Let me end this section by sketching a possible solution to the problem of performative self-contradiction. It is plausible to assume that there exist entirely unconscious encoding mechanisms in the brain, subpersonal mechanisms capable of forming unconscious memories even of absolutely prototypical MPE states, a registration of full-absorption episodes characterized not only by low, but by *minimal* complexity. It is also well known that the human mind quite frequently creates false memories (Roediger III & Marsh, 2009). Therefore, a retrieval of such unconscious memories referring to past selfless episodes into the level of conscious processing could conceivably lead to the creation of a specific type of false memory, by transforming them into an "autobiographical data-format", integrating them into a personal-level narrative. Sincere reports about selfless states could be therefore be sincere reports about false autobiographical memories. Strictly speaking, indexical linguistic expressions like "I", "this experience", "now", or "here" could never have been applied during the episode itself, but they are now used in sincerely referring to the content of one's memory. What makes these memories false is the introduction of spatiotemporal self-location and a subject/object dichotomy into the mental or linguistic representation of an episode that actually lacked these structural features at the time of its occurrence. Call this the "false-memory theory".

This point then leads us back to the first and second methodological difficulties exposed in this section:

- Autophenomenological reports about MPE are actually either reports about more complex experiential configurations or reports about false autobiographical memories;
- whenever they are presented as reports about full-absorption episodes or of prototypical MPE *as such* we must necessarily assume that they are *post-hoc* confabulations (or are at least strongly theory-contaminated on a subpersonal level);
- any *positive* phenomenological characterization of MPE, if based on first-person reports, will always originate from more complex states that have resulted from a superimposition of specific contents onto one's MPE.

4.6 Conclusion: MPE as the phenomenology of tonic alertness

4.6.1 Terminology: What is alertness?

The empirical literature is often unclear and ambiguous regarding the use of notions like "arousal", "vigilance", "sustained attention", "wakefulness", or "alertness". However, there are two semantic elements one finds in a large majority of treatments: First, *epistemic capacity*, second, absence or *lacking specificity* of representational content. For example, and full accordance with the proposal to be formulated below, Oken, Salinsky and Elas speak of "activation states of cerebral cortex that impact the ability to process information where the activation itself contains no specific information" (Oken, Salinsky, & Elsas, 2006, p. 1885). It is exactly the representation of such states which is relevant for MPE, but from a philosophical perspective the existing terminology to describe such states is highly inconsistent. As Oken and colleagues write:

> Unfortunately, no terms are ideal to describe these states of cortical activation since most terms are in broad use with varied associations and there are not perfect physiological markers. The term *vigilance*, in particular, has unfortunately been used in different ways by different groups of scientists. [...]
> The field has been hindered by inconsistent or poorly defined terminology. Researchers should be particularly careful about the usage of the term *vigilance*, understanding there are factors contributing to vigilance other than just arousal on the sleep–wake dimension. Avoidance of the term vigilance because of its varied definitions would be most helpful. The use of the term *tonic alertness* may be preferable since it does not already have a confusing history (Oken et al., 2006, 1885; 1991).

Let us, for the purposes of this paper, clearly distinguish between the physical, the functional, and the phenomenological levels of description. I will use terms as follows:

"Arousal":

- A graded **physical property** of the human brain.

    → A purely *physical boundary condition* determining the depth of information processing available to the organism as a whole.

    → A *vital parameter* that must be predictively controlled (e.g., to generate the sleep-wake cycle).

"Tonic alertness":

- A graded **functional property** determining the capacity sustained attention.

    → A causal function *resulting* from the successful control of arousal over longer periods of time, in the absence of an external cue.

    → A functional property causally enabling capacities like, orientation, executive control, attention, and epistemic agency on the mental level.

"Wakefulness":

- A graded **phenomenal property**, which is sometimes introspectively accessible.

    → A *representation* of tonic alertness.

→ The major component of the phenomenal character of MPE (**PC1**), the *primary dimension* of phenomenal state space.

What is the relationship between these three properties? In humans, the functional property of tonic alertness is *realized* by certain physical properties of the brain and we do not yet know what exactly these properties are. Conceivably, it could be realized by very different physical properties in machines, but for neurotypical humans it is highly plausible to assume that a critical level of cortical arousal is the most central necessary condition. If the right level of arousal as such is sufficient for tonic alertness is currently unknown, but unlikely.

The phenomenal experience of wakefulness is a *representation* or, perhaps, a *predictive Bayesian model* of the functional property of tonic alertness. Both the phenomenal property of "wakefulness" and the functional property of "tonic alertness" cannot exist without the brain. But tonic alertness, if it is not represented or internally modelled can exist without the conscious experience of wakefulness (as in some cases of unresponsive wakefulness syndrome, cf. section 7.2). Put differently, an organism can be alert without *knowing* that it is alert. It is important to understand that the representational level of description opens the possibility for misrepresentation, therefore some instances of MPE could be misrepresentations of tonic alertness. Plausibly, the phenomenology of wakefulness can exist in low-arousal states (as in lucid dreamless sleep [LDS], cf. section 5). This raises many interesting questions, for example if LDS would have to count as a *misrepresentation* of tonic alertness, if you will, as a hallucination or a dream *of* phenomenal wakefulness.

What do we know about the functional property connecting the physical and the phenomenological level? Tonic alertness is a global functional property, in part physically realized by the level of cortical arousal, and it is hypothesized to be one of the core functions of the operculo-cingular (CO) network (Sadaghiani & D'Esposito, 2015). It is one component of what we call the capacity for "attention" in folk-psychological contexts, and its function for the organism consists in "achieving and maintaining a state of high sensitivity to incoming stimuli" (Posner, 2008, p. 193). Our neuroscientific understanding of the CO-network is gradually improving, and it already provides us with important pointers that are consistent with our phenomenological analysis: Tonic alertness is a) a sustained, b) internally initiated "preparedness to process and respond" which implies a capacity for attentional agency, of being able to "co-engage" phasic alertness and selective attention directed at a specific task (Sadaghiani & D'Esposito, 2015, p. 2763). Just like the quality of mindfulness cultivated by contemplative traditions it can be described as a global and general mechanism "of keeping cognitive faculties available for current processing demands and holding unwanted activity at bay" (Sadaghiani & D'Esposito, 2015, p. 2764).

It is also empirically plausible to assume that tonic alertness is a central causally enabling factor in the realization of two other functional properties which are distinct, but closely related: *orientation* and *executive* control. Here, it is important to note that all three functions – alertness as such, orientation, executive control in terms of attentional agency, - are epistemic capacities, which, at a given point in time, a biological organism may have or not have. Please also note how the paradigmatic cases of "pure" consciousness discussed above are precisely characterized by an absence of spatiotemporal self-location (i.e., during full-

absorption episodes, the subject is not *oriented* towards time or space) and absence of attentional or cognitive agency (i.e., mental actions like *controlling* one's thought processes or the focus of attention). The current proposal says that phenomenality (i.e., a minimal form of conscious experience) emerges precisely whenever the first functional property – tonic alertness, the causal enabler of orientation and executive control - is probabilistically *represented*. In other words: Minimal phenomenal experience is a predictive model of a specific set of epistemic capacities, namely, "high stimulus-readiness", "orientation readiness", plus the mere "capacity for mental self-control" on the level of attention and cognition. As a Bayesian model is therefore closely related to the subjective confidence that these capacities will lead to the creation of internal states carrying epistemic value in the future. More simply, we could say that MPE is an expectation of knowledge.

4.6.2 The phenomenology of phasic alertness

We can now formulate the empirical hypothesis that what satisfies WAKEFULNESS on the phenomenological level is the existence of a *representation* of tonic alertness. However, on the functional level of description, it is important to distinguish tonic alertness itself, which is long-lasting and independent of external cues, from short episodes of heightened, but externally triggered arousal. In the words of Weinbach and Henik:

> Alertness is commonly divided into two different modes of function: tonic and phasic alertness. Tonic alertness, which is also known as "intrinsic alertness," designates the internal control of wakefulness or arousal in the absence of an external cue in a top-down manner (Sturm et al., 1999; Sturm & Willmes, 2001). Phasic alertness, on the other hand, is a short-lived effect of achieving high levels of alertness following a salient external event (Posner, 1978; Posner, 2008; Sturm & Willmes, 2001). (Posner, 1978; Posner, 2008; Sturm & Willmes, 2001). (Weinbach & Henik, 2012, p. 1530)

On a conceptual level, we can interpret tonic alertness as the epistemic capacity to create and sustain the sensitivity to incoming perceptual stimuli. This capacity, in turn, is causally related to the epistemic capacity for orientation: the creation of an integrated and egocentric representational space plus the epistemic capacity for executive control of attention. PC1, the phenomenology of wakefulness, is the conscious representation of the first epistemic capacity. Perspectivalness, the phenomenology of being an epistemic agent or a "knowing self", is the conscious representation of the second and third capacity.

How does this relate to the phenomenology of phasic alertness? A paradigm example is the rapid change in attention due to a brief event, like an unexpected sound in your environment. The phenomenology of phasic alertness is the experience of being forced to take a new attentional perspective, of re-orienting the "arrow of intentionality" (T. Metzinger, 2006, 2017). It is a brief episode in which the first-person perspective as consciously experienced is shifted to a new epistemic target. Scientifically, phasic alertness is typically studied with infrequent, unpredictable warning signals preceding the presentation of a stimulus, shortening reaction time. Whereas intrinsic, tonic alertness is the basis for operations such as orienting and selective attention, these operations themselves have a distinct phenomenal character of their own: The dynamic process of *relocating* attentional resources or suddenly shifting to a new target for executive control, typically following an experience of unexpected salience detection. What makes an experience salient depends on many different factors: Perceptual features like contrast, clarity, unexpected size or intensity clearly play a causal

role, but the overall emotional, motivational or cognitive context may be equally important in triggering the phenomenology of phasic alertness.

4.6.3 The phenomenology of tonic alertness

The phenomenology of tonic alertness is the conscious experience of wakefulness. It only fluctuates on the order of minutes to hours. It is entirely non-conceptual, not the result of an inference, or a judgement about the current behavioral state (whether one is asleep or awake). As such, the experience of bare wakefulness has no internal structure, but it has certain global aspects that can be introspectively discerned.

The phenomenology of wakefulness is the experience of *openness*: The organism is now sensitive to incoming stimuli, it can perceive the environment and internal states of its own body, it is *open to the world* and it represents this very fact. In this sense, MPE is a non-egoic, non-conceptual form of self-knowledge.

Second, the phenomenology of wakefulness is the experience of possessing a set of *epistemic capacities*: For example, the organism now has the capacity of orienting itself in time and space, of taking a perspective. It non-conceptually knows about this capacity but does not exert it. As research on the experience of "pure consciousness" in meditation demonstrates, the phenomenology of bare wakefulness is independent of time-representation and of self-location in a spatial frame of reference, it often appears without it. On the functional level of analysis, tonic alertness and the orientation reflex can be dissociated: Tonic alertness is what causally *enables* spatiotemporal self-location and orientation, but it is not identical with it. The conscious experience of wakefulness is a representation of exactly this epistemic capacity, the mere capacity to know where you are, at what point in time you exist, and how you are currently positioned in the order of things. The phenomenal experience of wakefulness is a way in which an organism can non-conceptually *know* about the existence of this mere capacity and this is also the reason why the relevant phenomenal character is extremely abstract and subtle. Epistemologically, therefore, wakefulness is an abstract, non-egoic form of self-knowledge. However, this knowledge appears under a specific inner mode of presentation, by using an internal representational format which does not yet involve subject-object structure, time-representation, or spatial embodiment. The ensuing phenomenology can therefore be described as a selfless, timeless, and non-spatial, as a *model of an epistemic space*, a space in which knowledge states can occur. Phenomenologically, this space is indivisible, it has no centre and no boundaries.[8]

Third, there is a quality of subjective confidence, an *expectation of epistemic states*. Computationally, wakefulness is a statistical hypothesis, a representation of the probability

---

[8] Again, one should avoid a phenomenological *non-sequitur* at this point: To say that the space of pure consciousness is "boundless" does not imply that there is an explicit experience of infinite expansion or of large distances. Rather, it means that there is no "other side beyond the boundary" to which attention could shift. To give an example, in visual awareness attention can shift from a red color patch into adjacent green color patch, transgressing a chromatic boundary. For the phenomenology of tonic alertness there is no such boundary, because everything outside of the space of wakefulness cannot be deliberately attended to. It is simply unconscious.

that veridical perception will actually occur. For example, it is part of the phenomenology of waking up from deep sleep that there is a *sense of confidence* that perceptual states will very soon occur, that not only is one now "open to the world", but that one can and will very soon know where one is and what time it is. Phenomenologically, the gradual transition to the wake state is also characterized by another subtle and intuitive presentiment, something one might perhaps term a phenomenal "foreshadowing of selfhood". This is an expectation of mental agency and the capacity for global self-control, the subjective, but as yet non-egoic confidence that a full-blown first-person perspective involving executive control on the mental level plus an extended autobiographical self-model will very soon appear. Therefore, the phenomenal quality of subjective confidence not only relates to the environment, but it can also extend to the likelihood of egoic self-knowledge, to the capacity of "predicting oneself into existence".

Finally, there is a fourth discernible aspect characterizing the phenomenology of bare, wakeful awareness, and it is related the absence of time-representation. Wakefulness has a quality of spontaneity, as it is ahistorical and contains not even an implicit representation of its own causal history. It is a spontaneous presence.

This gives us a first phenomenological description of the relevant target state: It is an experience of openness to the world, in combination with an abstract, non-conceptual representation of mere epistemic capacity, plus a domain-general sense of confidence, of the likelihood of knowledge, but as yet without object. The target state is experienced as unconstructed and spontaneously present. It is also interesting to note that *prima facie* the method of phenomenal contrast (Siegel, 2007, 135pp) cannot be applied to the phenomenology of tonic alertness: It seems impossible to point to a pair of two experiences which only differ in terms of the relevant content - if there is no phenomenology of wakefulness, then there simply is no conscious experience altogether. In other words, if prototypical MPE as characterized above really is consciousness *as such*, then it makes no sense to contrast its specific phenomenal character with what could only be unconscious states *devoid* of such character. If the above is correct, phenomenality is a global representation of subjective confidence, relative to the possibility of knowledge.

4.7 Summary

The primary goal of this section was to lay the first foundations for a minimal model explanation of consciousness (cf. section 2). One aim is to develop a new entry point for research, taking a closer look at a specific phenomenological subset, namely, "pure consciousness"-experiences. The first result was that such experiences clearly exist. A second result is that there are actually fewer paradigmatic reports of such experiences in the literature than previously assumed. Third, while such reports have strong common phenomenological denominators, they are always "multidimensional" in the sense that they express multiple features and not only one single, unequivocal form of phenomenal character. Fourth, a first analysis of the relevant literature clearly shows that developing a more fine-grained and empirically grounded phenomenology of "pure consciousness"-episodes is confronted with a number of serious methodological obstacles, for example the problem of embodied theory contamination (cf. section 4.2), the problem of a positive characterization of the relevant type of phenomenal character (cf. section 4.3), and the

problem of performative self-contradiction (cf. section 4.5). These problems seem to have affected and hindered scientific research in the past. As one early reviewer stated: *None of the studies reviewed (…) met the five criteria: adequate reliability and validity, a comprehensive phenomenological assessment, and control for introspective sensitization and demand characteristics* (Pekala, 1991, p. 78) .

On the other hand, we now have extracted six provisional phenomenological constraints for the concept of "minimal phenomenal experience" (MPE) in hand, which also provide us with a very first theoretical prototype. These constraints will be used to construct and evaluate a first database of relevant case studies (Appendix I) and to prepare neuroscientific studies by developing a novel psychometric instrument (Appendix II):

- **PC1**: WAKEFULNESS: the phenomenal quality of tonic alertness.
- **PC2**: LOW COMPLEXITY of reportable content: an absence of high-level symbolic mental content (i.e., conceptual or propositional thought or mind wandering), but also of perceptual, sensorimotor, of emotional content (as in full-absorption episodes).
- **PC3**: SELF-LUMINOSITY: a phenomenal property of MPE typically described as "radiance", "brilliance", or the "clear light" of primordial awareness.
- **PC4**: INTROSPECTIVE AVAILABILITY: we can sometimes actively direct introspective attention to MPE and we can distinguish possible states by the degree of actually ongoing access.
- **PC5**: EPISTEMICITY; as MPE is an *epistemic relation* ("awareness-of",) *if* MPE is successfully introspected, then we would predict a distinct phenomenal character of epistemicity or subjective *confidence*.
- **PC6**: TRANSPARENCY/OPACITY: like all other phenomenal representations, MPE can vary along a spectrum of opacity and transparency. It can be experienced *as* a representation or not.

As shown in figure 1, even these preliminary constraints already allow us to describe and compare specific classes of phenomenal states in a new and more fine-grained way:

	PC1 Wakefulness	PC2 Complexity of reportable content	PC3 Self-luminosity	PC4 Ongoing introspective access to PC1	PC5 Epistemicity	PC6 Transparency of PC1
Normal Waking	dark	dark	light	light	dark	dark
Effortless mindfulness	dark	medium	medium	dark	medium	medium
Full-absorption	dark	light	dark	light	light	light
Lucid dream sleep	dark	dark	light	medium	light	dark
Lucid dreamless sleep	dark	light	dark	light	light	light
Dream sleep	light	dark	light	light	medium	dark
Deep sleep	light	medium	light	light	medium	dark
Minimally conscious state	medium	medium	light	light	light	dark
Unresponsive wakefulness syndrome	light	light	light	light	light	light

Figure 1: The six phenomenological constraints of PC1 – PC6 as applied to waking, two types of contemplative states, four types of sleep states, and two coma-related types of state. Medium shading indicates that the degree of constraint satisfaction is variable.

To give a first example, we can now compare the two examples of "effortless mindfulness" and "full absorption" discussed in section 4.2 to normal wake states in terms of the complexity of reportable representational content:
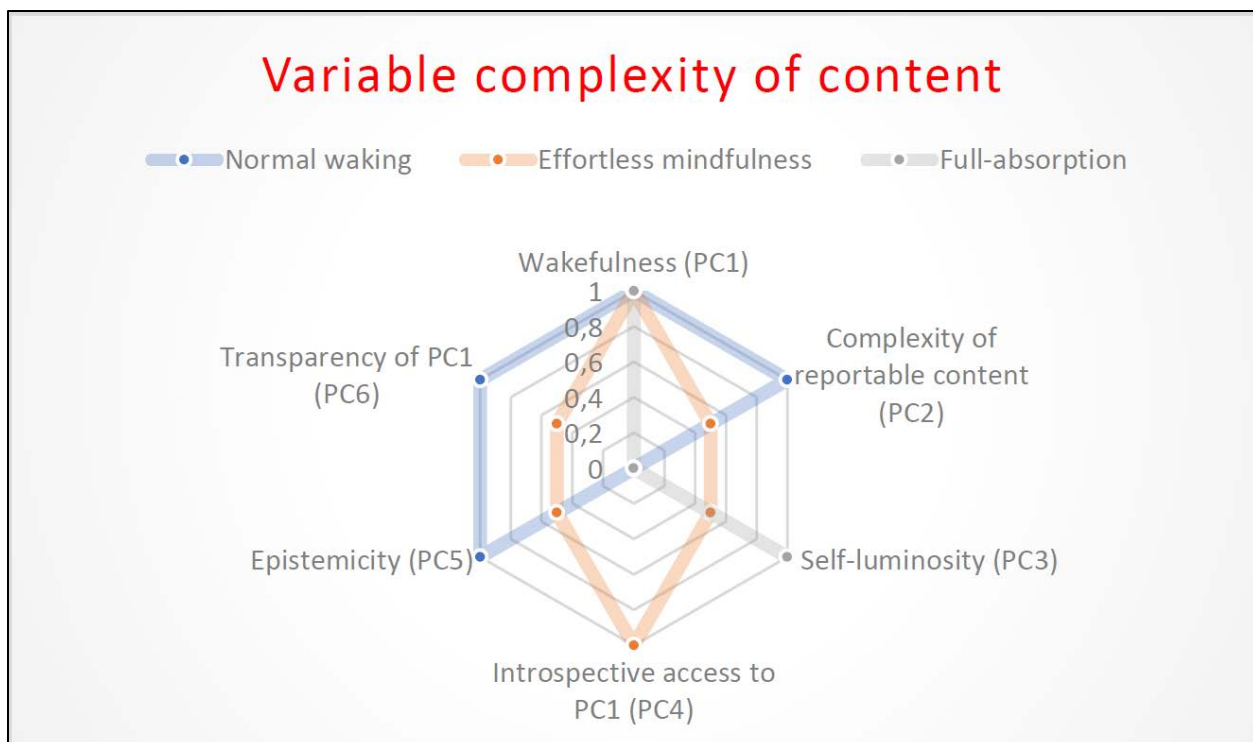
Figure 2: The complexity of reportable content of effortless mindfulness as lying between prototypical MPE (ineffable) and normal waking (which adds cognitive agency and spontaneous, task-unrelated thought).

This new taxonomy also yields a first systematic insight. If we look at rows 3 and 5 in figure 1, then we discover that they are the only two rows presenting us with identical descriptions for two different phenomenal state classes: The *prototypical* example of MPE can be found in deep, fully absorbed contemplative states as well as in lucid dreamless sleep.

5. Lucid Dreamless Sleep: The instantiation of a minimal model in the real world?

It is now natural to ask if we find credible reports about anything coming close to an "instantiated minimal model" of conscious experience in humankind's phenomenological traditions. What would be a state in which MPE was maximally expressed because the superimposed content was minimal? Is there perhaps a class of conscious states that comes close to the "common phenomenological denominator" mentioned in section 3.2, the members of which exhibit the highest degree of prototypicality and which therefore strongly correlate to the sought-after *minimal* theoretical model of conscious experience itself? Naturally occurring states with low autonomous mental activity, a representation of tonic alertness plus a high degree of functional disconnection from the physical body's sensory and motor systems would be obvious candidates. Dreamless deep sleep is such a state and given the minimal model approach plus the set of constraints developed above, there is a strong theoretical motivation to investigate MPE during dreamless deep sleep.

5.1 What is lucid dreamless sleep?

"Lucid dreamless sleep experience" (LDS) is a distinct phenomenal state-class that has been anecdotally reported for many centuries. It has played a role in classical Indian philosophy of mind (Advaita-Vedanta). It is often also called "witnessing sleep" (see section 5.2). During episodes of LDS, an inner experience of bare awareness accompanies dreamless deep sleep. It

is mostly reported by highly advanced practitioners of meditation. The phenomenology of LDS satisfies PC1 (WAKEFULNESS), PC2 (LOW COMPLEXITY), and sometimes PC3 (SELF-LUMINOSITY). Given the existence of reports about LDS (cf. Appendix I for examples), it is also plausible that PC4 (Moorean INTROSPECTIVE AVAILABILITY) is at least sometimes satisfied, though the superimposition principle of course holds.

What makes LDS interesting as a real-world candidate for minimal phenomenal experience is not only independent theoretical motivation, but also recent empirical evidence for the occurrence of phenomenal states during periods of dreamless deep sleep (Dentico et al., 2016; Ferrarelli et al., 2013; Maruthai et al., 2016; Mason et al., 1997; Mason & Orme-Johnson, 2010; Jennifer M. Windt, Nielsen, & Thompson, 2016). Evan Thompson (see also Thompson, 2015b, 265pp), Jennifer Windt, and Tore Nielsen have made a major contribution in this area by re-directing attention to the question of phenomenal experience during dreamless sleep (Thompson, 2015b, 2015; Jennifer M. Windt, 2015; Jennifer M. Windt et al., 2016). Thompson writes:

> [...] we need to distinguish between ordinary dreamless sleep and lucid dreamless sleep. Ordinary dreamless sleep is the sleep of ignorance, in which awareness is described as being in total darkness. Lucid dreamless sleep is described as a state in which awareness is luminous and without an object (free of thoughts and images). Whereas lucid dreaming consists in knowing that you are dreaming, lucid dreamless sleep is said to consist in being able to witness the state of dreamless sleep and recall its phenomenal clarity upon waking up. Although the background metaphysics of Yoga, Vedānta, and Indo-Tibetan Buddhism differ in significant ways, they all describe lucid dreamless sleep as disclosing a basal level of pre-personal consciousness that lies deeper than the modes of awareness that characterize the ego-centred waking and dreaming states. (Thompson, 2015, p. 15)

Jennifer Windt introduced the concept of "minimal phenomenal experience" (Jennifer M. Windt, 2015). One theoretical motivation was the question of what, conceptually, one would have to subtract from a previously existing minimal model of *self*-consciousness (MPS, or "minimal phenomenal selfhood" as introduced in Blanke & T. Metzinger, 2009) in order to arrive at a minimal form of consciousness *as such*. For Windt, the step from MPS to MPE is the step of subtracting the two conditions of a weak first-person perspective and spatial self-location:

> The key idea is that while even the simplest forms of dreaming are characterized by phenomenal selfhood, or the experience of being or having a self, the transition from dreaming to dreamless sleep experience occurs when even this minimal form of phenomenal selfhood is lost. While the analysis of dreaming can help identify the conditions for minimal phenomenal selfhood, the analysis of dreamless sleep experience may provide a glimpse of an even simpler (and perhaps even minimal) form of phenomenal experience. (Jennifer M. Windt, 2015, p. 3)

A merit of Windt's approach is that it offers a positive model of the transition from MPS to MPE and offers the first explicit theory of what MPE really is – namely, something resembling an almost empty, but merely *temporal* frame of reference, envisioning dreamless sleep experience as involving "pure subjective temporality, or phenomenal experience characterized only by the phenomenal *now* and the sense of duration, but devoid of any further intentional content" (Jennifer M. Windt, 2015, p. 1). Hence, she writes:

> The moment at which self-location dissolves—or at which minimal phenomenal selfhood is replaced with the maximum unit of identification—involves a transition to the type of pure

> subjective temporality that earlier, I suggested might be the phenomenal mark of dreamless sleep experience. As lucid dreaming gives way to lucid dreamless sleep experience, minimal phenomenal selfhood shades into pure phenomenality, in which phenomenal experience is characterized only by its temporal structure. (Jennifer M. Windt, 2015, p. 22)

I agree that for phenomenal experience *as such,* neither explicit embodiment, minimal spatial self-location, nor epistemic agency are necessary conditions, and that LDS is the best concrete candidate for the conceptual "essence" of conscious experience we currently have. However, Windt's MPE-theory has three major shortcomings:

- **Contentlessness**: The phenomenologically explicit nature of the "emptiness" and/or the *concreteness* of the "mode-neutral brightness" (SELF-LUMINOSITY) reported is not made fully intelligible by *abstract* ideas of an empty temporal frame of reference nor by the notion of "pure temporality".
- **Atemporality**: As it stands, the proposal ignores the phenomenology of "timelessness" as the conscious experience of presence and the sense of duration are clearly forms of explicit temporal experience (LOW COMPLEXITY). Many of the case studies presented above an in Appendix I seem to show that there is simply *no* time-representation in MPE.
- EPISTEMICITY: Windt's model ignores the "phenomenal signature of knowing" (Thomas Metzinger & Windt, 2014; Thomas Metzinger & Jennifer M. Windt, 2015; section 3.1). The fact that there is an additional phenomenal quality of "epistemicity" or "passive, non-agentive knowing" is something that every theory of MPE should explain (PC5). It finds its clearest expression in traditional terms like "empty cognizance" or "witnessing sleep".

5.2 What is "witness-consciousness"?

The notion of "witness-consciousness" refers to a term in classical Advaita Vedanta philosophy (*sākṣin*; Sanskrit: साक्षी). Just like "turiya", "rigpa", "samadhi", or "dharmakāya" (already mentioned above) it is yet another example of phenomenological notions with a long tradition in Indian philosophy, quite obviously bearing relevance to the present sketch of an interdisciplinary research program. The earliest appearance of the idea seems to be found in the Atharva Veda, in the Muṇḍaka Upaniṣad (Chatterjee, 1982, p. 339), whereas the source of the philosophical concept "sākṣin" is obscure (Fort, 1984, p. 287, n. 2; Gupta, 1998,

p. 19).Witness-consciousness is that which makes all knowledge possible, cannot itself become an object of knowledge, and is self-luminous. Bina Gupta gives the following epistemological characteristics:

1. The witness consciousness, although the basis of all knowing, is different from the object known. It is implied in every act of knowing. It is the ultimate subject; it can never become an object of knowledge.

2. It is the pure element of awareness in all knowing. It is one, immutable, indivisible reality.

3. It shines by its own light; it is self-luminous.

4. It is different from the empirical individual (*jīva*; Sanskrit: जीव), who cognizes and enjoys. In other words, it is different from the empirical individual who is caught up in the triple states of waking, dreaming, and dreamless sleep. (Gupta, 1998, p. 18)

On a purely phenomenological reading that abstracts away from all metaphysical and epistemological aspects of the term "witness-consciousness", we find something that strongly

resembles MPE. It is a self-luminous form of bare awareness which can be superimposed by intentional objects during single "acts of knowing", a non-egoic type of phenomenal character which has nothing to do with any phenomenal self-representation as a cognitive agent or an entity possessing subjective preferences, and, perhaps most importantly, something that could in principle be present in waking states, during dreaming, and in dreamless sleep, simply because it is *different* from any phenomenal self-model only portraying the "empirical individual" (Fort, 1980). In short, witness-consciousness is an intrinsically non-reactive, non-egoic, and non-conceptual experience of knowing. One interesting reading is of the *sākṣin* as a "field" of consciousness, within which the distinction of subject, object, and the process of knowing arises as a "context for contents, or 'space' in which forms may appear" (Fort, 1984, p. 278).

Albahari defines witness-consciousness as a mode-neutral knowing or awareness with intrinsic phenomenal character (Albahari, 2009, p. 66), although it remains unclear what exactly the non-relational "intrinsicality" of witness-consciousness really amounts to and what exactly the specific phenomenal character of interest consists in. Wolfgang Fasching interestingly describes contemplative attempts to cultivate witness-consciousness as a process of "de-superimposition":

> Instead of identifying certain configurations of *experienced contents* as being 'oneself', one begins to experience oneself as the abiding *experiencing itself* (the taking place of *presence*) of any contents. De-superimposition means radically distinguishing oneself from all objects by no longer delimiting oneself (as an 'inside') as opposed to the objects 'out there'. […] In the 'de-identified' mode of experiencing that is strived for, one completely lets go of 'oneself' and becomes nothing but 'seeing', without any distinct 'seer' standing apart from the 'seen'. (Fasching, 2011, pp. 211–212)

This point very clearly relates to **PC5** (the EPISTEMICITY constraint) and the tentative interpretation of Moore's concept of consciousness as an aperspectival and non-agentive form of meta-awareness which is co-instantiated with all forms of conscious knowledge. If, however, MPE were co-instantiated with a state of dreamless deep sleep (if you will, a naturally occurring form of complete "de-superimposition"), then we would expect exactly what the classical term "witnessing sleep" seemingly refers to: An aperspectival, selfless, and contentless process of knowing *as such*, an isolated phenomenology of pure observation. If this reasoning is correct, then any minimal model of conscious experience will have to accommodate the fact that, quite obviously, all the preliminary constraints extracted from the literature in order to develop a first working concept of MPE in section 2 can sometimes be satisfied during dreamless deep sleep. Perhaps most interestingly, the "phenomenal signature of knowing" (PC5; Thomas Metzinger & Windt, 2014; Thomas Metzinger & Jennifer M. Windt, 2015) and the phenomenology of tonic alertness (WAKEFULNESS) could then appear during a physiological state that is normally described as fully *unconscious*, something our minimal model of conscious experience would have to make intelligible on a theoretical level. In the next section, I will present such a model.

6. The ARAS-model theory of minimal phenomenal experience

The ARAS-model theory of minimal phenomenal experience claims that MPE is the representational content of an ARAS-model, because MPE dynamically represents a specific subset of the activity of the *Ascending Reticular Activation System* (ARAS). Again, please note

how "ARAS" here only functions as a placeholder for a much more fine-grained and detailed empirical story to be told in the future. The general idea is that MPE is a predictive, global, top-down model of the strongest internal signal source the cortex has to deal with. What we call "bare awareness" or "consciousness as such" today is a dynamic, integrated representation of the very bottom-up process by which the system activates itself. Once again, the following passages are just a first attempt to explain the general neurocomputational idea, they are *not* meant as a speculative kind of "armchair functional neuroanatomy". The general idea, however, is that the ARAS is the hidden causal core of the strongest *internal* signal source the brain deals with, which it must "explain away", regulate and control. The ARAS-model theory of MPE states that minimal phenomenal experience is a predictive model of the ARAS-signal, which is the most central causally enabling factor in the emergence of conscious experience.

6.1 What is the *Ascending Reticular Formation* (ARAS)?

The reticular activating system is a well-documented network in the brain stem, which arouses and activates the cerebral cortex. It is a central part of the internal mechanism controlling alertness as well as the most important circadian rhythm of the human organism: the sleep-wake cycle. Its ascending, widely distributed fibres send signals to the cortex, regulating the level of arousal. The ARAS can be described as a functional system of auto-activation (M. Steriade, 1996a); it is the system by which the brain wakes itself up (M. Steriade, 1996b; Mircea Steriade & McCarley, 2013).
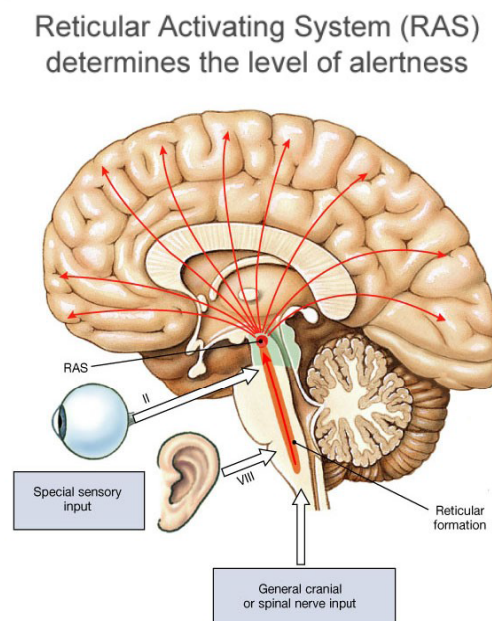

Figure 3: The Ascending Reticular Activation System (ARAS). Figure courtesy of slideshare.net. For an artificial brain simulation illustrating the process of auto-activation, see https://www.youtube.com/watch?v=acdL-8wZhLc.

6.2. The need to control the process of auto-activation

A first point frequently overlooked in the philosophical literature is that the brain's level of arousal – just like body temperature, glucose levels, or blood oxygenation – is a central *vital parameter*, a physical property that any human organism must successfully control in order stay alert and to regain consciousness after deep sleep.

The Yerkes-Dodson-law (Yerkes & Dodson, 1908) states an empirical relationship between arousal and performance: For most tasks, performance increases with physiological arousal, but only up to a point. When levels of arousal become too high, performance decreases. Therefore, cortical arousal is an important physical parameter in the brain that must be regulated and controlled to achieve optimal performance on a functional level.
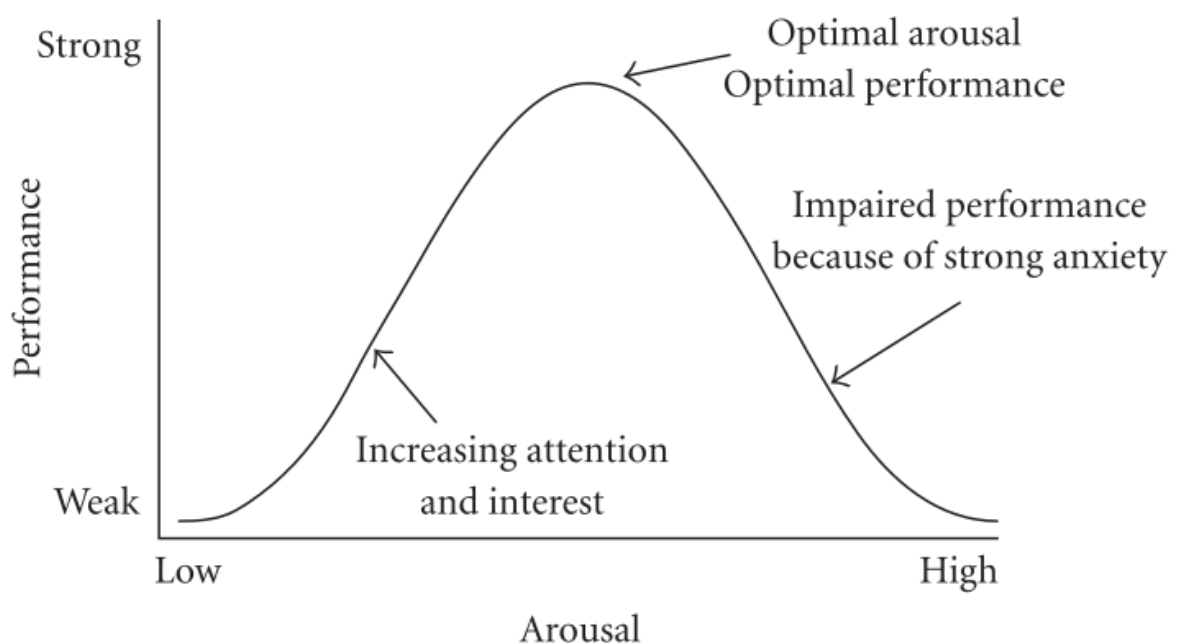
Figure 4: The Yerkes-Dodson-law shows how optimal performance presupposes successful control of the level of arousal (source: HebbianYerkesDodson, CC0 1.0).

6.3 Epistemic and non-epistemic dimensions of conscious experience

In empirical research on consciousness it has been commonplace to distinguish two main dimensions characterizing phenomenal states, namely "arousal" and "awareness" (T. Bachmann & Hudetz, 2014). "Arousal" refers to the physical degree of arousal in the brain, whereas "awareness" typically is used to describe the complexity of reportable content. It is important to understand that "arousal" is a non-epistemic dimension, while "awareness" is generally interpreted in a quasi-Moorean way as "awareness-of". Awareness means that the organism non-conceptually *represents* the world and itself in a certain way, and often such representations create epistemic states making them a form of knowledge. On the other hand, representation implies the possibility of misrepresentation (F. Dretske, 1986). In the figure below, we can see how different types of states are currently categorized according to

these two dimensions, for example in coma research. Attentive readers will notice that one logical possibility is not depicted in the figure below.
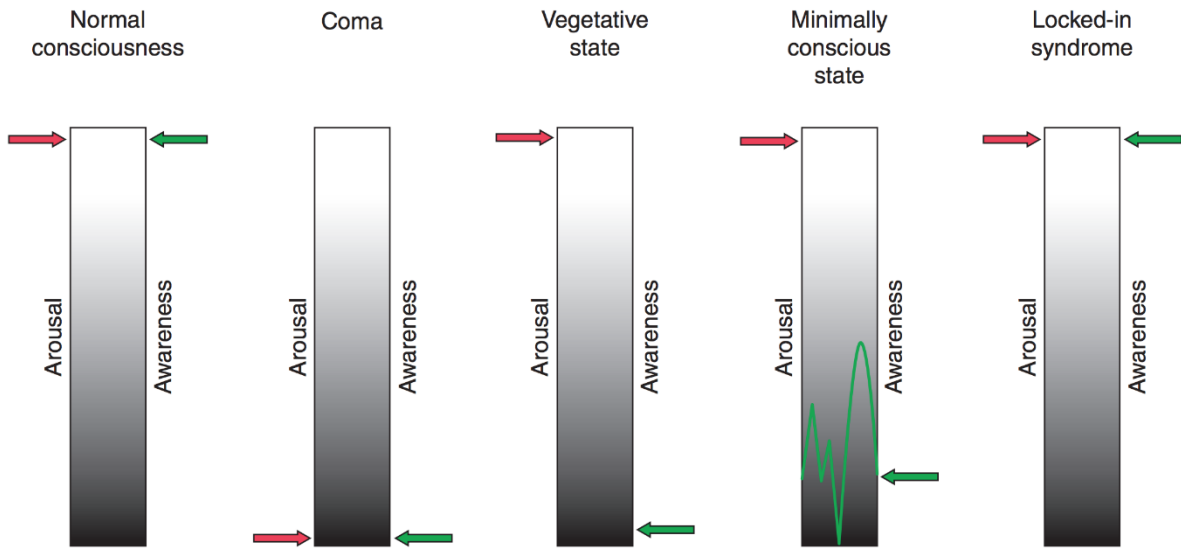

Figure 5: Arousal and awareness as non-epistemic and epistemic dimensions across different neurophenomenological configurations. What is missing is the possibility of a high-awareness state enduring during low-arousal configurations like deep sleep, for example in "lucid dreamless sleep", cf. figure 6 below and figure 11 in section 7.7. Figure 5 courtesy of Steve Laureys (S. Laureys, Boly, Moonen, & Maquet, 2009).
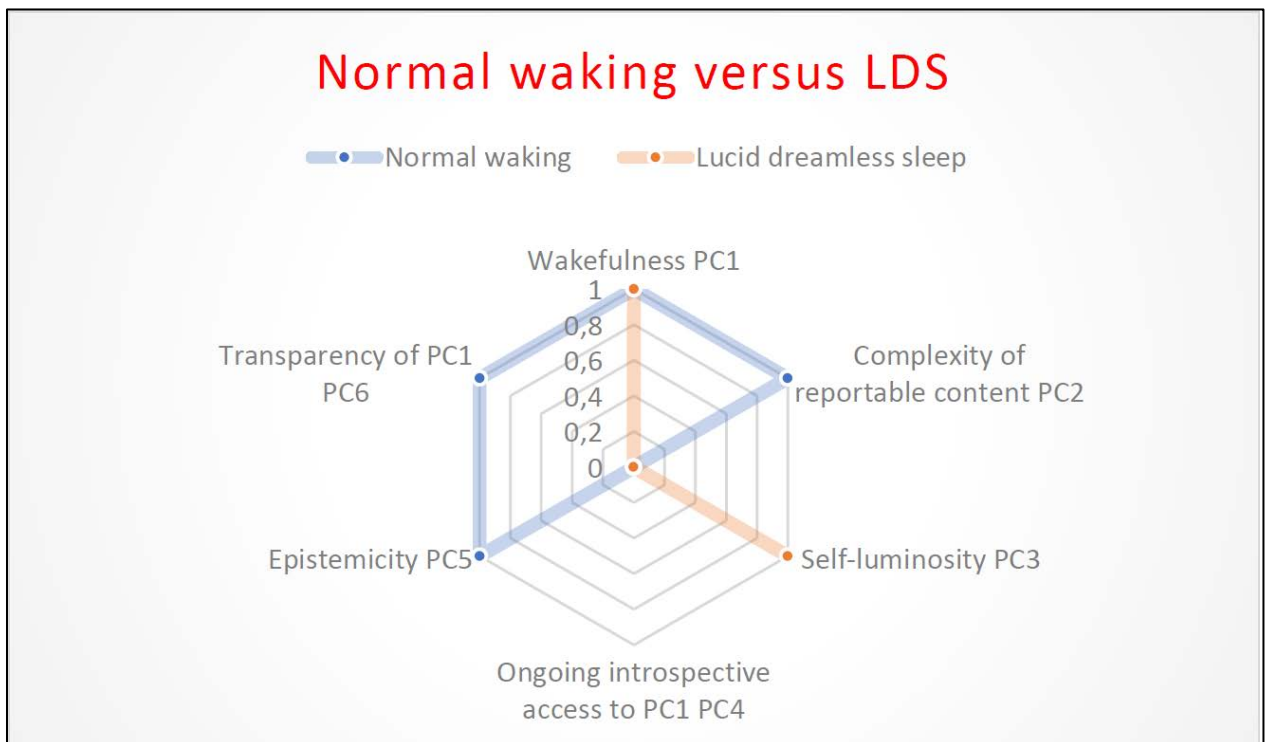
Figure 6: Lucid dreamless sleep fully satisfies only WAKEFULNESS and SSELF-LUMINOSITY, while normal waking satisfies WAKEFULNESS, LOW COMPLEXITY , EPISTEMICITY, and TRANSPARENCY.

Here is one possibility of arranging different phenomenological categories in the two-dimensional plane created by the two dimensions of "wakefulness" (which here only refers to functional alertness; see the terminology introduced in section 4.6.1) and awareness:
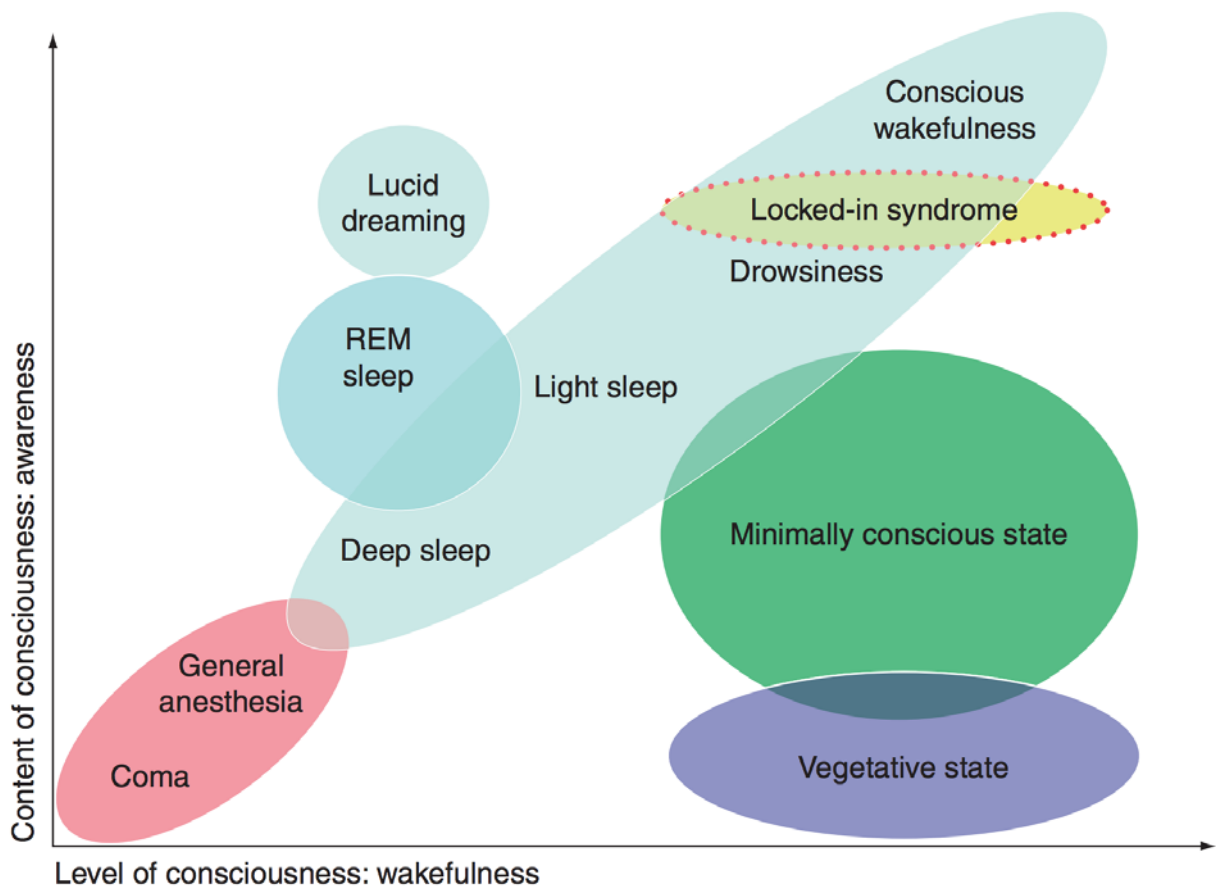

Figure 7: A well-known two-dimensional taxonomy of conscious states. "Wakefulness" on the abscissa clearly refers to the physical property of cortical arousal, as conceptually distinguished from the *phenomenal* property of wakefulness at the end of section 4.6.1 above. "Awareness" on the ordinate refers to the complexity of reportable conscious content, not to the degree of phenomenality as such. This makes it impossible to distinguish phenomenally experienced wakefulness either from complex representational content or from mere physical arousal. The special case of an awareness *of* one's own "wakefulness" (i.e., the having an internal representation of the physical property of arousal and/or the functional property of tonic alertness) would be hard to depict. MPE, the representation of tonic alertness, involves low complexity for phenomenal content (Low Complexity; a small value on the ordinate as here used) and a high value of *consciously experienced* Wakefulness (PC1). Prototypical MPE would be above general anaesthesia, ordinary deep sleep, and the vegetative state in terms of content complexity, but below the minimally conscious state. LDS (lucid dreamless sleep) would be *between* the vegetative and the minimally conscious states in terms of phenomenal content, but the value for physical arousal on the abscissa would at least be as high as that of the vegetative state, minimally conscious state, locked-in syndrome, or ordinary wake states (see also Bayne et al., 2016, p. 411). Figure courtesy of Steve Laureys.
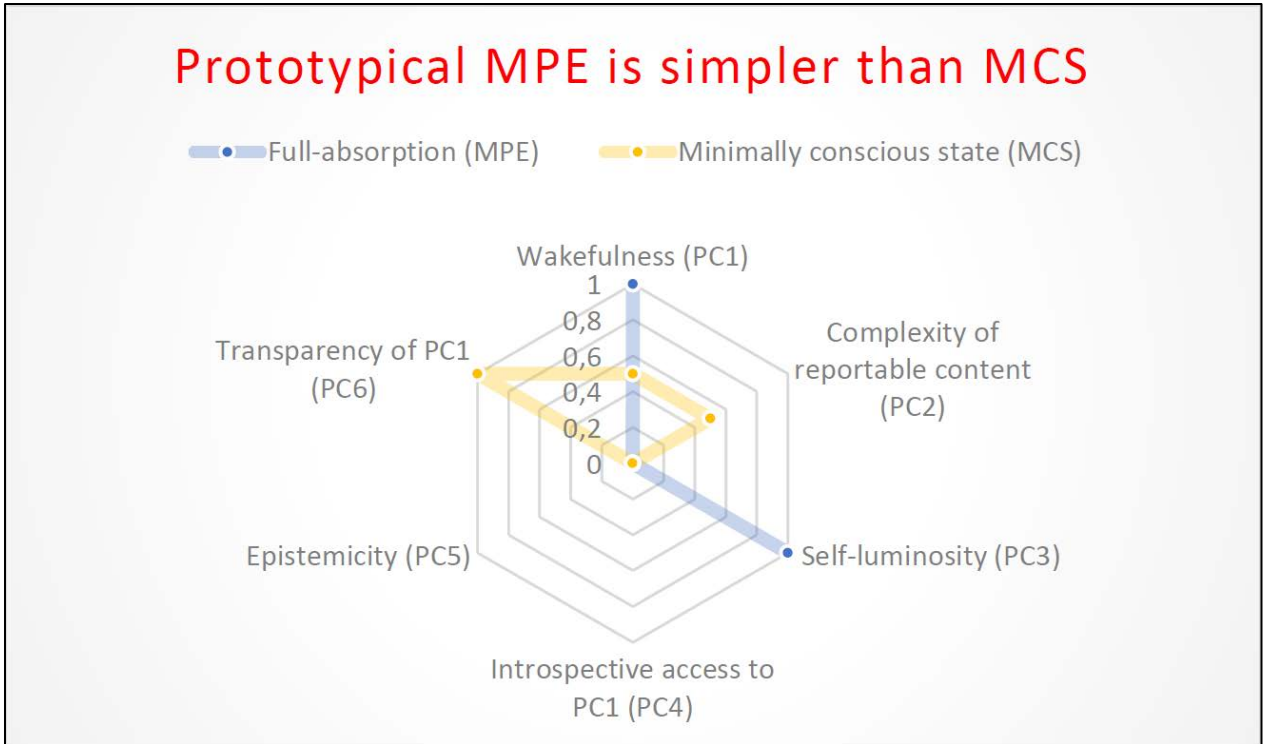
Figure 8: MCS satisfies three phenomenological constraints, MPE only two.

Figure 9: A comparison between prototypical MPE, MCS, and the (unconscious) UWS-state.

A first, rather intuitive, step in the reasoning behind the ARAS-model of conscious experience is that - if there really exists a most general, abstract and potentially basic form of "contentless" phenomenal character could be strongly correlated with the most general and basic internal signal source the brain must control and its attempt to "understand" it. Under a predictive-processing approach (Wiese & Metzinger, 2017) "understanding" could be equivalent to "explaining away" by creating a generative internal model that helps the cortex to cancel out this signal. A second step in the argument is that the cortical level of arousal is a central physical parameter the organism must control. It is empirically plausible to assume an evolutionary pressure towards optimizing this specific kind of internal self-control. The third step is to integrate the epistemic and the non-epistemic aspects of conscious experience on a conceptual level: Here, the idea is that for a *minimal* model of consciousness we might take a closer look at these two dimensions and ask: Is there any evidence for an internal representation *of* arousal or tonic alertness as such? The answer is yes.

We can interestingly (and hypothetically) analyse MPE as a dynamic representation of the current level of arousal and as a predictive model used for controlling the widely distributed ARAS-signal (see section 6.3). In other words, although MPE is traditionally described as a "pure" form of conscious experience, as a non-intentional "mere emptiness", if we look at it from a third-person perspective it actually *has* an epistemic dimension and an intentional object: MPE is an internal model of the current level of auto-activation. The non-representational[9], purely functional bottom-up process of auto-activation has triggered a top-down-model of itself. This model is what has, in the past, been called "pure consciousness" or "awareness as such"; it is what embodies the phenomenal character of wakefulness (see section 4.6.1)

In the next section, I will sketch a coarse-grained predictive processing model of MPE. However, before doing so let me recapitulate the major stages of the reasoning that has led from the minimal model approach to this claim.

    P1 The strongest and most general signal source the human brain must regulate and control is the activity of the *Ascending Reticular Activation System* (ARAS).
        →   Therefore, a logical next step is to investigate the relation of this signal source to our hypothetical, most general and all-pervading form of conscious content, namely, MPE.
    P2 The ARAS-signal itself is non-representational; it is a purely causal form of auto-activation.
        →   If there were an internal representation of this signal, it would necessarily be a representation of something *contentless*.
        →   The phenomenal character of such a representation, if there was one, could not be contrasted or compared with a situation in which it was absent, because the act of "contrasting", necessarily, presupposes exactly those epistemic capacities only causally enabled by sustained tonic alertness.
    P3 At any point in time, there exists an optimal level of arousal.
        →   Yerkes-Dodson-law.

---

[9] The ARAS-signal does not *represent* in any stronger sense e.g., Fred Dretske (1983), but it certainly carries *information* about a number of other hidden causes, for example blood flow, additional organismic variables, or the status of the current daylight cycle. Environmental factors and a wider causal/statistical context do exist, but the ARAS-signal as such does not explicitly represent them.

C1 There will have been an evolutionary pressure to adaptively control ARAS-activity to continuously achieve the optimal level of arousal in a flexible and context-sensitive manner.
C2 Whenever this form of biological self-control takes place, we can describe it as involving a form of explicit, but non-egoic self-knowledge, an inner representation of the hidden cause of the ARAS-signal within the organism itself.

Here is some further background reasoning from a more conceptual perspective:

Moore is right:

- Phenomenologically, "awareness-of" is an epistemic relation.
  - → EPISTEMICITY: There is a distinct phenomenal character of "knowingness" or subjective confidence in MPE as introspected.

However, according to the present proposal Moore is wrong on some other points:

- Consciousness *as such* is not a 2nd-order epistemic state.
  - → The ARAS-model theory of consciousness says that MPE is a global, first-order epistemic state representing a global, but *non-intentional* state (the non-epistemic, contentless ARAS-signal).
  - → The ARAS-model theory of consciousness says that a (convoluted, reportable) 2nd-order epistemic state is only created by activating an epistemic agent model (against M6 and M7).

This gives us a new perspective on the ideas by G. E. Moore and Jennifer Windt, which were discussed above:

- Phenomenological reports show that during full-absorption episodes subjects are not oriented in time at all, there is no experience of "nowness", duration, or abstract temporal structure (as in the notion of "pure temporality").
  - → The ARAS-model theory of consciousness implies that MPE precedes whatever mechanisms in the brain are responsible for time-representation and orientation within a temporal frame of reference.
- The original model of MPE misses the phenomenal character of EPISTEMICITY (PC5).
  - → The ARAS-model theory of consciousness says that phenomenal epistemicity is subjective confidence, which can in turn be analysed as introspectively accessed *model-evidence*.[10]
- Windt's original model does not yet help us to understand the concreteness of the phenomenal character of Moorean "mere emptiness".
  - → The ARAS-model theory of consciousness says that MPE (if introspected) *must* appear as "mere emptiness" because it models a hidden causal force, a non-intentional vehicle property, in an abstract manner.

6.4 Towards a predictive processing model of MPE

This section offers a first sketch of an idealized and parsimonious computational model of MPE. We have already seen that the level of cortical arousal is a vital parameter, and that it is plausible to assume that the human organism needs to control it in an optimal manner. However, in trying to understand how auto-activation leads to auto-illumination, we must never forget that an ARAS-model can be described on many different levels of description. For example, we can analyse it under a control-theoretic, a cybernetic, and a predictive processing perspective at the same time (Seth, 2015). This reveals that ARAS-modelling can

---

[10] See note 5 for details of the idea and Meyniel et al. (2015)

be viewed as a form of *self-control* with the help of a generative model. The natural conclusion is that MPE might result whenever the brain uses a generative model to predict and control the ARAS-signal.

Just as perceptual models can be said to predict sensory signals and to "cancel out" such signals by generating a statistical hypothesis about their occurrence (a "mock signal" or "virtual version" of the expected input), we can now point out that it is highly plausible that the massive, distributed, but *non-perceptual* ARAS-signal will sometimes also be the target of such an internal modelling strategy. In this case the result would not simply be a top-down prediction of some sensory input in the sense of classical perceptual inference. Rather, it would then become what I would term "virtual auto-activation": An all-pervading, top-down prediction about the level of arousal to be expected in the next instant, a "mock version" of the ARAS-signal if you will. Consciousness, at its deepest and most minimal level, would then be a virtual model of the highly distributed functional process of auto-activation.

However, in the special case of the human ARAS-model, some further aspects will be interestingly different. First, we would have a maximal causal proximity between model and target, leading to an especially deep and robust form of biological self-representation. What the model represents is the hidden cause of the stimulus. However, this time, the hidden cause is not a signal source in the external or interoceptive (Seth, 2013) environment of the brain. Rather, the ARAS is a part of the central nervous system itself, a fact that contributes to a unique situation:

- Biological nervous systems represent the world by predicting own internal states via hierarchical Bayesian modelling, thereby inferring the causes of sensory stimuli causing events on sensory sheets.
- The neural body has no self-directed receptor system; it is perceptually blind to itself.
- The ARAS-model is the centrally relevant case in which the brain represents and predictively controls a *low-level physical state* of the neural body itself (namely, the hidden cause of cortical arousal) by creating another brain state (namely, the ARAS-model).
- Therefore, MPE is a non-perceptual, mode-neutral way in which the brain "sees itself". Put differently, in its earliest origin consciousness is the functional solution to the problem that the brain is perceptually blind to itself but nevertheless has to understand the process by which it activates itself.

Computation in the brain crucially involves an interplay between top-down and bottom-up processing, and PP emphasizes the relative weighting of top-down and bottom-up signals in both perception and action. MPE is a top-down model of a fundamental, domain-general process of bottom-up activation, which is created when bottom-up signals of the ARAS-system have been sufficiently re-weighted to exert a stronger causal influence on cortical processing systems (as in the process of waking up from deep sleep).It is not a form of interoceptive inference (Seth, 2013), because no receptor systems or visceromotor, hormonal or immunological predictions are involved (Barrett & Simmons, 2015), but it bears a structural resemblance to the early "activation synthesis" model of dreaming (A. Hobson & McCarley, 1977; J. Hobson, Pace-Schott, & Stickgold, 2000; Jennifer Windt, 2015, 59pp).

Please recall how in section 4.6.1 we termed the functional-level correlate of the conscious experience of wakefulness "tonic alertness". MPE can now be described as an "expectation of

tonic alertness": An estimation of the degree of sustained, global activation the system can expect in the future. Standardly, estimates can be regarded as statistical hypotheses, which serve to explain sensory signals. The ARAS-model explains away an *internal, non-perceptual signal* by generating a statistical hypothesis about expected levels of arousal. It also induces an evidentiary boundary, a statistical envelope or shell secluding conscious cortical processing from the subcortical causes of arousal (Hohwy, 2016, p. 264). The ARAS-signal itself is continuous, an analogue signal, which comes in degrees and variable strengths, but the ARAS-model, is a discrete, individual entity delimited by a Markov-blanket. Any given organism either has MPE or it has not, it either has a distinct ARAS-model or it has not. Technically speaking what impinges on the Markov-blanket will still have to be called a "sensory" signal, because any relation connecting a hidden cause to a model of this hidden cause and its context has "sensory" states. But this is a merely terminological point: Here, the states predicted have nothing to do with sensory perception and we are not dealing with any form of direct perceptual inference. It follows that, if any phenomenology results from this process, it would have to generate a new kind of phenomenal character *sui generis*. This phenomenal character would be concrete, but nevertheless non-perceptual, non-motor, and non-cognitive and as it precedes the construction of spatial and temporal frames of reference it would be devoid of spatial or temporal forms of conscious content.

PP also involves computing estimates of random variables and prediction error minimization. In computing prediction errors, prediction error terms must be weighted by precision estimates, and a central goal of PP is to *minimize* precision-weighted prediction errors. For the special case of internal auto-activation, prediction error will generally be low and model-evidence will be high because a stable ARAS-signal is always present whenever conscious experience occurs. During ordinary waking states, phasic alertness will of course vary according to the present situation and unexpected environmental stimuli impinging on the organism's sensory organs. Nevertheless, *tonic* alertness remains relatively stable, at least in comparison to many other sensory variables. Strong prediction errors may certainly occur, for example when we are in danger of fainting or when we are exhausted. Generally, however, the ARAS-signal can be expected to be one of the most stable, permanent, and invariant input sources the brain is confronted with.

Under normal conditions, the precision of the top-down prediction of the top-down ARAS-model is so high that only a very strong prediction error (for example, when we suddenly faint or unexpectedly lose consciousness in another way) can change it and thereby attract introspective attention. If, however, we learn to volitionally reduce model precision, the bottom-up prediction error caused by the ARAS-signal can be turned into an explicit content of conscious experience. First, MPE was transparent, now it is opaque. This may be what happens in some forms of contemplative parctive.

 PP additionally enables predictive control (Seth, 2015) and is action-oriented in the sense that the organism can act to change its sensory input to fit with its predictions and thereby minimize prediction error. Among other benefits, this enables the organism to regulate its vital parameters (like levels of blood oxygenation, blood sugar, etc.). Here, the important point is that the cortical level of arousal is a prime example of such a vital parameter that needs to be continuously regulated. This point opens an interesting perspective: What exactly

is "active inference" for cortical arousal and the ARAS-model, and what exactly are the inner and outer "actions" by which human beings change the ARAS-input to fit with their current "wakefulness prediction" and to minimize prediction error?
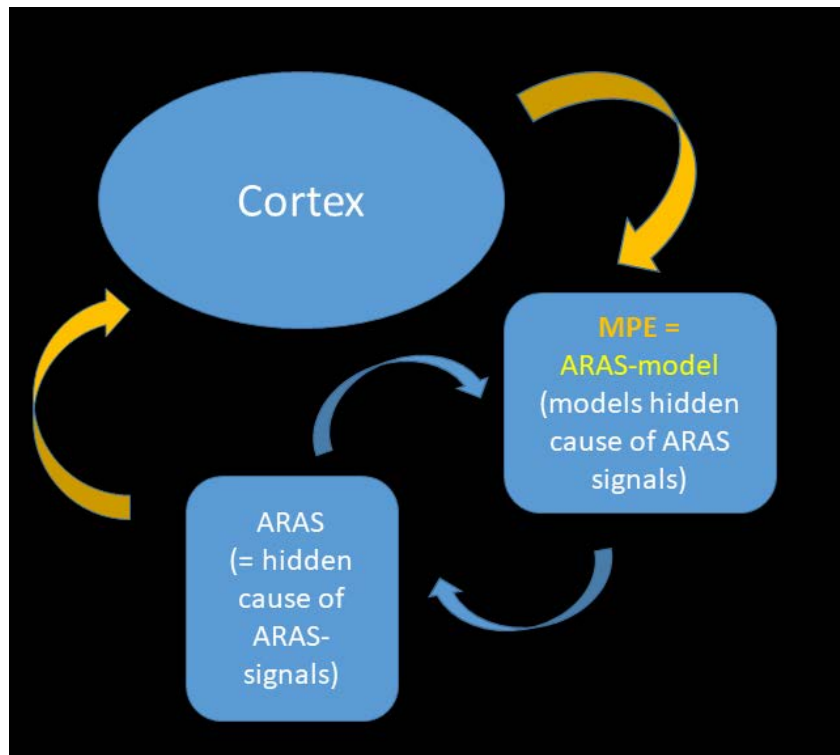

Figure 10: Whenever the ARAS-signal has reached sufficient intensity, the cortex reacts by creating an ARAS-model, trying to extract the hidden cause in its internal environment. During episodes of conscious processing, the ARAS-model can be used to predictively control the ARAS signal, thereby optimizing the cortical level of arousal. As this is an abstraction, please note that the ARAS-model could itself be implemented in the cortex.

What about INTROSPECTIVE AVAILABILITY, introspective access to MPE? The first Moorean constraint says that MPE can sometimes be available for introspective attention. Under the PP-approach, attention can be analysed as the process of optimizing precision estimates. This is interesting in the context of advanced meditators who can now be described in a completely new way, namely as people who seek to systematically cultivate and stabilize their ARAS-model, optimizing precision expectations for the ARAS-signal. One may speculate that highly advanced meditators and subjects experiencing episodes of lucid dreamless sleep have trained, developed and stabilized a stand-alone model of tonic alertness. Functionally, it is highly plausible to assume that repeatedly and systematically attending to the internal model of the ARAS-signal improves precision expectations and stabilizes the level of arousal and tonic alertness by stabilizing the ARAS-model. Phenomenologically, the prediction would be that this form of mental practice intermittently creates an introspective EAM directed at the property of wakefulness *as such*. But again, from a computational perspective, the ARAS-signal itself does not directly represent any internal or external events. Again, to avoid any misunderstandings at this point: The ARAS-signal does not *represent* in any philosophically more interesting sense (e.g., Fred Dretske, 1983), but it certainly carries *information* about a number of other hidden causes (say, the current daylight cycle). There are environmental factors, and they form a wider context. The signal causing internal auto-activation, however,

has a variable signal strength and statistical structure in time, which can be predicted and internally modelled.

If we are interested in the prototypical phenomenological case of "pure" consciousness as such, it is important to understand how the ARAS-signal *precedes* all the higher cortical functions – the extraction of modality-specific sensory features, object formation, scene and event segmentation, time-representation, self-modelling, and so on – which it causally enables. Therefore, perhaps the most natural way the resulting phenomenology *can* be described is as "gazing into the void", as wakefully attending to a functionally autonomous, basic, and self-luminous form of emptiness (cf. the list of negative characterizations presented in section 4). If so, this would be fully consistent with the Tibetan Buddhist definition of *rig pa* (pure consciousness) as "knowledge of the ground". With a minimal computational model in hand, let us look at some testable empirical predictions the ARAS-model theory of conscious experience makes.

7. Empirical predictions

7.1 Coma

- If the ARAS-signal completely disappears, then all phenomenal experience vanishes.

This prediction is almost trivial, and has been confirmed by modern coma science (Steven Laureys et al., 2004). Arousal is a necessary condition for conscious experience to occur.

7.2 Unresponsive wakefulness syndrome

- If ARAS-input exists *without* an ARAS-model, unresponsive wakefulness syndrome will result.

In states of high arousal without model-based predictive control, there will be no conscious awareness of the environment and no phenomenal self-model (T. Metzinger, 2003a). However, many of the *functional* properties defining wakefulness will be preserved, such as an intact sleep-wake-cycle (Steven Laureys et al., 2010). Please recall the distinction between physical property of arousal, the functional property of tonic alertness, and the phenomenal property of wakefulness at introduced in section 4.6.1 and compare figure 9.

7.3. Mindfulness practice and tonic alertness

- Systematically cultivating introspective access to the ARAS-model should functionally promote intrinsic, tonic alertness and the phenomenology of WAKEFULNESS (PC1).

As might be expected, classical mindfulness practice engenders increased and sustained wakefulness and gradually makes the relevant states more and more effortless (i.e., it removes the superimposition by the subjective experience of agency). It is plausible that advanced practitioners succeed in constructing a stand-alone ARAS-model, preserving the phenomenology of wakefulness during periods of dreamless deep sleep (cf. section 4). Much more research is needed, but a start has been made (Britton, Lindahl, Cahn, Davis, & Goldman, 2014; Lutz et al., 2009).

7.4 Emptiness

- In empirical studies, the statistics of phenomenological reports referring to MPE should show it as contentless and empty (PC2; LOW COMPLEXITY).

As explained above, prototypical MPE *as such* is strictly ineffable because the present model predicts either amnesia or misrepresentation on the level of autobiographic memory (cf. section 4.4). However, weakly superimposed states can of course be targeted by neurophenomenological research programs yielding valuable data, though not about the relevant phenomenal state "in itself". Empirical studies can therefore yield valuable insights about the statistical landscape of autophenomenological reports in relation to neural correlates, as for example created by systematic experience sampling in combination with neuroimaging studies and other methods (see for example Garrison et al., 2013; Schoenberg, Ruf, Churchill, Brown, & Brewer, 2018). The phenomenal correlate of predictively controlled ARAS-input should be described by subjects as contentless and devoid of internal structure, like an indivisible space lacking boundaries and a centre (a "mere emptiness", exactly as Stace, Moore, or Buddhist philosophy say).

Again, the ARAS-signal is not a *sensory* signal in the sense that it activates a specific type of receptor or perceptual module, a physically realized transductor. Therefore, its model should not generate any form of *sensory* phenomenal character. There is no "what it is like" in terms of a distinct subjective quality corresponding to the specific data-format generated by a specific type of physical transductor, like the introspectively detectable "visual", "auditory", or "visceral" character in genuine perceptual experience. From the brain's perspective, there is almost nothing in its previously known conscious models of reality to which the system could compare it. The ARAS-signal is the central causal element in a purely functional process of auto-activation. Hence it is *non-representational;* it has no intentional object.

The same is not true of the ARAS-*model*. From a third-person perspective, we can say that the ARAS-model has representational content and that there actually is an intentional object: The current level of cortical arousal as caused by ARAS-activity, perhaps the expected precision of the ARAS-signal itself (see section 7.2). But it is now clear why the ARAS-model, if introspected, *must* be reported as contentless: There are no sensory features, there are no multimodal objects, and there is no scene that has been segmented in time or space, the projective geometry characterizing most other global representations is absent (Williford et al., 2018). Viewed from a personal level, there is nothing it is *like*, because it cannot be compared to any known experience (Nagel, 1974). On the personal level, no conscious contrast class is available (Siegel, 2007). As it is the earliest and most fundamental level of conscious model-construction, phenomenological reports should describe it as non-egoic: There should be no self-location in time, no self-location in space, and no localized and transparent UI or unit of identification (for a definition and references, see T. Metzinger, 2018).

Prototypical MPE is an internal model of an epistemic space which has not yet been fragmented by a subject/object dichotomy or an agentive arrow of attention (cf. section 4.6.3). The empirical prediction is that in the statistics of phenomenological reports we will also find no sense of bodily agency, no EAM, no phenomenal *self-as-subject* and no internal first-person perspective. Therefore, the corresponding phenomenological profile should be "non-dual" in the sense that – besides "qualia" or irreducible sensory primitives - it also lacks the second characteristic Thomas Nagel attempted to isolate as the hallmark of subjectivity,

namely, an individual first-person perspective (Nagel, 1986; Chapter 4).

7.5 Fundamentality

- As the causal influence of the ARAS-signal is maximally distributed and causally enables *all* forms of conscious processing, and as the ARAS-*model* represents its hidden cause, the corresponding phenomenology should be ubiquitous and all-pervading, implicitly underlying all other forms of phenomenal character.

Here, the prediction could be that MPE is implicitly present in all forms of phenomenal character, but mostly characterized by full transparency, therefore leading to an unnoticed absence. There is a number of arguments that could support this traditional philosophical intuition. First, in the large majority of cultural and educational contexts there simply is no *expectation* that there is a distinct phenomenal character of MPE that could form the object of introspective attention. No one ever tells us that there is a phenomenology of tonic alertness we could attend to. To quote Moore, one needs to "know that there is something to look for" (Moore, 1903, p. 450). Second, we are shallow introspectors. For more than two thirds of their conscious life-time, human beings are not at all able to control their introspective attention, for example during periods of non-lucid dreaming or during spontaneous, task unrelated thought, as in daydreaming or mind wandering (T. Metzinger, 2015a). Third, to actively make the "contentless wakefulness behind everything" an explicit element of conscious experience requires the successful control of introspective attention, which thereby automatically superimposes the target state with content, for example with specific expectations, goal-directedness, and an attentional EAM. (This is the "Refrigerator Light Problem" again.)

However, on an empirical level subscribing to the fundamentality constraint would make the current theory extremely vulnerable and very easy to falsify: If there exists only a single type of phenomenal experience – say, during anaesthesia, unresponsive wakefulness syndrome (UWS), or NREM-sleep – which can be sustained in the absence of an ARAS-model, then having such a model is not a necessary condition for consciousness. Please also recall that I have not endorsed the fundamentality constraint (which ultimately originates in ancient philosophical notions like *turiya*) but have set it aside as an open empirical question. Trying to satisfy the fundamentality constraint leads to a number of very strong and highly specific empirical predictions:

- For example, in terms of prenatal phenomenology, one would conclude that the *first* episode of conscious experience emerging in the central nervous system of every human foetus must always be MPE: As the ARAS-model is the very first computational layer that has to come online before more complex states can emerge, the prediction is that the very first episode in any human being's conscious biography will inevitably be an experience of pure awareness. This is of course extremely speculative, but perhaps some future technology could make the issue experimentally tractable.
- If the fundamentality constraint can be validated, then an ARAS-model should also be present in ordinary, non-lucid states of dreaming. Here, one may argue that the dream self certainly *thinks* that it is awake and that because phasic alertness exists in the

dream state, the phenomenology of being awake actually is present. But as there is no attentional control during the dream state, the non-lucid dream self can never direct its attention to the quality of tonic alertness. It may actually turn out that it is the absence of functional property of tonic alertness itself, which characterizes the dream state, that only phasic alertness is present in the non-lucid dreams state (section 4.6.2).

- However, a third prediction would be that an ARAS model and a more or less explicit state of MPE also underlie not only selfless states like lucid dreamless sleep (LDS), but also states belonging to the *other* two phenomenological categories of dreamless sleep experience proposed by Windt, Nielsen, and Thompson (2016). That is, MPE should also form the phenomenological background of isolated bodily sensations or perceptual experiences during dreamless sleep.

- In addition, MPE should also be implicitly present in isolated or static visuospatial, auditory, or kinaesthetic imagery lacking a clear hallucinatory context, movement sensations, and propositional thought (sleep thinking), which all count as dreamless experiences (cf. BOX 1 in Jennifer M. Windt et al., 2016).

- A fifth highly specific prediction is that recovering coma patients on a path from unresponsive wakefulness syndrome (UWS) to the minimally conscious state (MCS) should briefly pass through a stage of MPE, perhaps resembling lucid dreamless sleep (see sections 4 and 7.6 below). The statistics of retrospective autophenomenological reports of such patients should therefore reflect this transition, another clearly testable prediction.

- Sixth, in healthy subjects experiencing the ordinary process of waking up, there should be a brief phase in which only the ARAS-model comes online. In other words, MPE should be an objectively detectable functional marker of global state transitions, which, however, is only rarely noticed by neurotypical subjects. Again, carefully designed phenomenological questionnaires should be able to isolate a statistically significant element here, corresponding to the earliest phase of the ordinary process of waking up.

7.6 Witness-consciousness

- The internal statistical model of the ARAS-signal has the property of integrated likelihood ("model evidence" in Bayesian statistics). Whenever introspectively detected, the prediction is that it should generate the phenomenal "signature of knowing", i.e., a specific phenomenal character of EPISTEMICITY, a reportable experience of subjective *confidence without content* in subjects' reports.

For complex, but low-superimposition states one can expect introspective reports referring to a phenomenology of aperspectival, all-pervading, and selfless signature of knowing, for example, not as if the reporting person was experiencing, but as if the world itself was "looking" (Ataria et al., 2015, p. 141).

7.7 Lucid dreamless sleep

- If a stand-alone ARAS-model were active during dreamless sleep, the statistics of autophenomenological reports should reflect the phenomenology of "witnessing sleep" *(or jagrat-sushupti in traditional terminology).*

Whenever tonic alertness is preserved in and represented during dreamless sleep, the expected phenomenal correlate would be wakeful, passive epistemicity without content. However, if the current theory is correct, the phenomenology really is not devoid of all content, but rather a first-order awareness of tonic alertness itself.
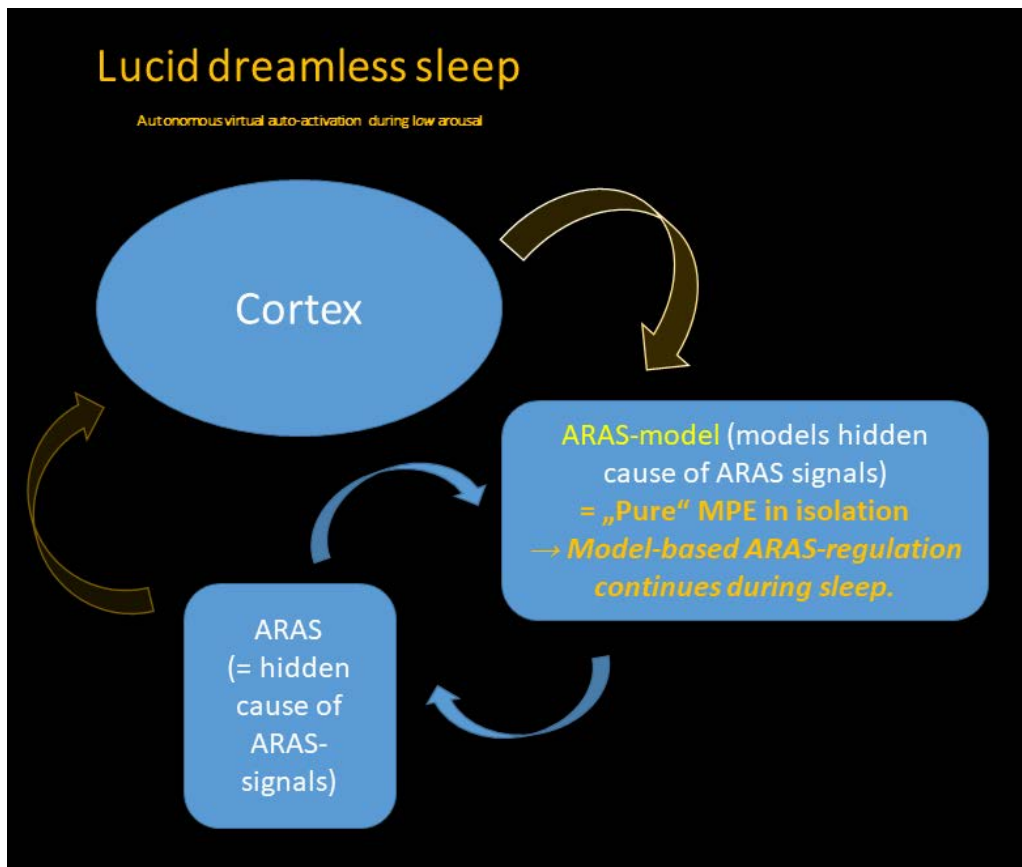

Figure 11: During deep sleep, the ARAS is in a different state, because bottom-up signals to the cortex have been down-regulated or re-weighted. If control and precision optimization have been trained for many thousands of hours (for example, in experienced practitioners of meditation), a stand-alone ARAS-model can be *sustained* in the absence of a strong ignition loop. See also figure 6 above, and please note that the model (as opposed to what the figure might suggest) could certainly be *physically realized* in the cortex.

LDS is the functional mirror image of UWS (unresponsive wakefulness syndrome), in the sense that it presents us with a state of low arousal combined with a state of high awareness. This leads to a second empirical prediction:

- Some UWS-patients *recovering* from a waking coma and on a trajectory to the minimally conscious state (MPS; Giacino et al., 2002; Wislowska, del Giudice, & Schabus, 2014) should pass through a stage of MPE in isolation.

This point leads to an interesting conceptual conclusion: The minimally conscious state is *not* the simplest state of conscious experience human beings can live through. The MCS is defined by signs of sustained awareness of the environment and oneself, for example as exemplified by purposeful behaviour as in reaching for, holding, or touching objects, simple vocalization, pursuit eye movement, gestures, or affective behaviours like appropriate smiling or crying (Giacino et al., 2002, p. 351). During lucid dreamless sleep (LDS) none of these defining features is present. Therefore, LDS and what has been termed "full-absorption episodes", and

not MCS, really are the simplest currently known form of conscious experience (cf. figure 8 for an illustration).

7.8 Maximization of the UI

- If an explicit ARAS-model were active during waking states, and if the UI were shifted from the EAM and/or the bodily self-model, this would create the classical phenomenology of "non-dual awareness".

The UI or "unit of identification" at a given point in time is whatever phenomenal character would lead to reports of the type "I *am* this!" (for a definition and references, see T. Metzinger, 2018). The UI is what an organism consciously *identifies* with. If an explicit and strongly de-superimposed form of MPE were to become the UI of a human being, then the prediction is that the result will be a global phenomenology of becoming identical with the process of conscious experience itself (cf. A-I @@).[11]

8. Conclusion

8.1. Towards a minimal model of conscious experience *as such*

We can now see that the ARAS-model theory of consciousness exactly fulfils the criteria for a minimal model outlined in section 2:

- It is a model of conscious experience that includes only the **core causal factors** giving rise to the target phenomenon, namely, the internal process of auto-activation by which the brain literally "wakes itself up" (M. Steriade, 1996b; Mircea Steriade & McCarley, 2013);
- it includes only those causal factors that make a decisive difference to the **actual** occurrence of the phenomenon and the **essential** phenomenal character itself (as expressed by phenomenological constraints **PC1, PC2, and PC**5);
- it is an idealized model of **universal and repeatable** features underlying all forms of conscious experience, even if the contents are highly disparate.

Before taking a second look at the provisional list of phenomenological constraints, let us briefly answer the questions posed in section 3 above:

8.2 Preliminary answers

Question 1:

---

[11] It is tempting to ask if MPE could be interpreted as an *epistemic modality*. To have a neurally realized predictive model in one's brain certainly constitutes a form of knowledge, and to have an ARAS-model is to be in an epistemic state. But there are very different modes of knowing, and some of them may not involve a subject/object relation. Let us assume an organism currently possesses an ARAS-model and that, phenomenologically, this model functions as its unit of identification. The model expresses a probability distribution and it predicts the level of arousal, continuously generating statistical hypotheses about the ARAS-signal. We can now say that the organism has acquired a highly specific kind of self-knowledge (namely, about certain physical and functional properties of its own nervous system), but that it internally presents this knowledge in a *phenomenal mode of identification*. It is knowledge by identification in that, on the level of phenomenal representation, the subject knows something specifically in virtue of *being* it. Cf. Forman (1993, p. 726); Merrell-Wolff (1973, pp. 93–97).

Is there a form of consciousness that does *not* instantiate MPS (i.e., a minimal form of selfhood, as in (Blanke & T. Metzinger, 2009) or an EAM (i.e., a minimal form of perspectivalness, as in (T. Metzinger, 2017)) or (Williford et al., 2018), see also note 7)?

- Yes, and it is called MPE.
- However, prototypical MPE in isolation is strictly ineffable.
- MPE is non-dual in that it is not internally structured by a self/world-prior, and most sociocultural contexts today clearly lack strong *cultural* priors for "pure" or "non-dual" forms of conscious experiences. We do not expect such states, and we do not teach our children to expect such experiences. Therefore, such priors will not be embodied on the level of individual brains. This leads to interesting questions about the social origin and cultural evolution of MPE-phenomenology: Are there differences in the internal modelling of tonic arousal throughout history, or relative to ethnic groups or religious traditions? Can one make predictions about the "optimal" sociocultural context that would support the appearance of this specific phenomenology?
    - If, as pointed out in section 4.1., the human self-model functions as the functional link between sociocultural and subpersonal levels of biological self-representation (Frith & Thomas Metzinger, 2016, pp. 209–210), then there is a deeper problem: A "non-dual cultural prior", even if it locally existed, cannot be embodied through the *self*-model, and this fact may even affect autobiographical memory. It is therefore possible that human beings instantiate MPE during deep sleep or during global state transitions like waking up but almost never report it.
    - Sincere autophenomenological reports either refer to false autobiographical memories or episodes of introspective attention directed at an ARAS-model. Therefore, such superimposed states will minimally add the phenomenology of attentional agency and of self-location in a temporal frame of reference: What is represented is a present moment, a dual state in which a knowing self is actively directed at "pure" consciousness.

Question 2:
Does pure consciousness exist?
- No, "pure" consciousness is not pure, because even minimal phenomenality *has* an intentional object: The expected level of arousal; the probability of tonic alertness.
- What has previously been called "pure consciousness" is actually an epistemic process; it is a dynamical, non-conceptual form of self-knowledge,[12] physically realized by a running ARAS-model.
- It can be attended to, but as such it is not reflexive or metarepresentational. Rather, it is a first-order probabilistic representation directed at a physical process in the human brain.
- One may speculate that systematically attending to it "ignites" a larger representational structure which is reflexive in the sense of continuously monitoring, tracking, or predicting itself over time. Call this the "self-ignited reflexivity hypothesis" (see next section).
- We now also have an explanation of why MPE necessarily appears as "empty" or "contentless" to the experiencing subject:
    → Although the functional process underlying wakefulness causally enables the activation of all other forms of phenomenal experience, it is evasive and, if made

---

[12] A philosophically interesting point, which for reasons of space I will not elaborate here, is that the ARAS-model is a deep causal layer of the human self-model. It clearly is a form of non-conceptual self-knowledge possessed by an organism, while at the same time it is not part of the *phenomenal* self-model and does not function as the UI in neurotypical humans.

explicit, its essential phenomenal character cannot be compared to any other form of conscious content the system knows.
→ There is no *phenomenal contrast* (Siegel, 2007), because the contrast class contains only unconscious functional states (e.g., UWS).
→ This absence of any potential "what it is like" (Nagel, 1974) leads to a misdescription of tonic alertness as "empty".

Question 3:
Can we develop a minimal model for phenomenality?
- Yes, there is an implicit kind of phenomenal character *sui generis* which can at times be *made* explicit and which underlies *all* other forms of phenomenal experience. It can be described on different levels of analysis, for example:
  → the conscious experience of WAKEFULNESS; i.e., a representation of mere *epistemic capacity*;
  → EPISTEMICITY a subjective, introspectively available *confidence* that "something can and will be known"; i.e., the consciously experienced degree of belief or "certitude" resulting from a non-conceptual representations Bayesian probability that new epistemic states will actually occur in the future;
  → a representation of the functional availability of an aperspectival and empty representational space; i.e., an inner, non-propositional knowledge of mere *representation*al capacity.

Question 4:
Is there one maximally abstract phenomenal property?
- Yes, there is a relevant form of phenomenal content: expected tonic alertness *as such*.
  → MPE is an internally controlled hallucination of tonic alertness.
  → MPE creates a distinct form of phenomenal character *sui generis*.
  → The experience characterized by WAKEFULNESS and EPISTEMICITY instantiates the most abstract phenomenal property, it subsumes or permeates all other forms of conscious contents but is mostly unnoticed and rarely introspected.

8.3. Preliminary constraints revisited

In closing, we can now take a second look at our six phenomenological constraints. WAKEFULNESS is satisfied by MPE-theory, nothing more needs to be said at this point. PC2 (lack of intentional content) is hypothetically rejected from the third-person perspective, as MPE actually *has* intentional content. But we also have an explanation as to why it should either be unnoticed or frequently misdescribed as "contentless" from a first-person perspective. There is no conceivable phenomenal contrast, because the contrast class would consist of entirely unconscious states: A truly minimal pair of situations in which there is either the experience of bare wakefulness or *no* experience of wakefulness at all would be like comparing a meditative state to coma - there would be no difference in the phenomenal character of overall experience, because coma is not an experience at all. Accordingly, there is also no mental operation which could allow us to compare MPE to an unconscious global state "in our mind's eye" as it were. But there is also no mental operation that allows us to juxtapose and compare MPE to the content of local conscious states, say, an individual colour experience. G. E. Moore saw this phenomenological point when (cf. M5) he said that no philosopher so far has been "able to hold *it* and *blue* before their minds and to compare them, in the same way in which they can compare *blue* and *green*" (Moore, 1903, p. 450).

Conversely, PC3 (Sᴇʟꜰ-ʟᴜᴍɪɴᴏꜱɪᴛʏ) is clearly the phenomenological constraint posing the greatest conceptual and intuitive difficulty. It refers to a phenomenal aspect of MPE traditionally described as "radiance", "brilliance", or as the "clear light" of primordial awareness. We can now offer two possible and more charitable interpretations:

- As an **epistemological metaphor**: Please note how, on a very literal reading, a "clear light" is something one cannot see. Clear light is not white light (which one could visually experience), and it also is not any other form of coloured light or visual experience. Clear light is a transcendental *condition for the possibility of experience* in our dominant sensory modality, it is what creates visibility in the first place. Thus, one charitable reading of the classical sources is to read "clarity" either as a visual metaphor for the phenomenal aspect of "mere epistemic capacity" characterizing MPE, or even more simply as an epistemological metaphor for the transcendental, experience-enabling function of the ARAS-model underlying MPE. In both cases, the metaphor makes use of a visual analogy, referring to our dominant sensory modality.

- As a **vehicle property**: If MPE is transparent, it will lead to a phenomenology of an immediately given epistemic space, of contentless wakefulness devoid of internal structure, something evasive and introspectively hard to detect. However, if we assume that highly trained introspective experts can learn how to penetrate the deeper construction process of the ARAS-model itself, if it should be true that advanced practitioners of meditation can actually learn to attend to earlier processing stages giving rise to an ARAS-model, then one would expect that MPE can at times become opaque. Here, the phenomenological prediction would be that even minimal phenomenal experience *itself* would acquire a quality of virtuality, of being a representational process lacking inherent existence.
The representational vehicle can actually be described as possessing *Eigenstrahlkraft* or self-luminosity, namely, in the sense that – viewed as a carrier of conscious content - it would realize a functionally autonomous kind of phenomenal experience, one that could in principle exist all by itself because it was phenomenologically self-subsistent.
In addition, the ARAS-signal itself clearly "radiates" in the sense that it spreads out spatially and activates many different brain areas at the same time. Furthermore, if it is possible to attend not to the model but also to the signal itself, then we might interpret the phenomenological concept of "brilliance" as a mode-neutral form of brightness, i.e. as an inner representation of mere stimulus strength. Again, we would have a visual metaphor – "brilliance" or "brightness" – applied to something incomparable to any other form of conscious experience. Therefore, the traditional concept of "self-luminosity" would be an attempt to pick out at least one aspect of MPE and creating an analogy by trying to say *what it is like* in the classical Nagelian sense already mentioned above (Nagel, 1974). If the experiential "self-luminosity" is not an epistemological metaphor, it could therefore simply refer to the sheer physical strength or "brightness" of the ARAS-signal itself.
This leaves us with the second option for a charitable interpretation: If MPE is

transparent, it represents a possibility – mere epistemic capacity, the possibility of knowing – as a directly given, mind-independent reality. If MPE is opaque, it represents a vehicle property, namely the physical "luminance" of the ARAS-signal.

PC4 (INTROSPECTIVE AVAILABILITY) is a functional property of the ARAS-model creating MPE, which can lead to more complex, superimposed states and which is sometimes realized and sometimes not. The conclusion is that an ARAS-model can sometimes become available as an object of introspective attention, although most of the time it isn't. We could even go so far as to postulate a mild form of "MPE-neglect" in neurotypical human beings, a deficit in attention to and awareness of the particular region of phenomenal space defined by tonic alertness – there was no evolutionary advantage of making tonic alertness a permanent, explicit part of the organism's model of reality, accordingly no cultural priors evolved.

On the other hand, under an evolutionary perspective, it is now plausible that many non-human animals have an ARAS-model and enjoy MPE, but only rarely make it the object of self-directed attention. A second conclusion would be that a major enabling condition for higher-order, introspective awareness of MPE would be a *prediction*, an expectation in the statistical sense or a culturally engrained belief that something to be found really exists. A this point it is interesting to recall what G. E. Moore said about consciousness as such, more than a century ago: "Yet it can be distinguished if we look attentively enough, and if we know that there is something to look for." (Moore, 1903, p. 450) There is also a strong resonance with the much older tradition of "pointing out" in Tibetan Buddhism (see the quotations presented above and in Appendix I, section 2.4).

PC5 (the subjective quality of "objectless EPISTEMICITY") is another essential aspect of the relevant phenomenal character which can sometimes result if MPE is successfully introspected. It leads to an experience of subjective *confidence*, a non-directed, non-agentive and aperspectival "knowingness", and one speculative proposal would be that *what* is consciously represented in such states is a specific property of the ARAS-model, namely, what appears as the abstract property of "model evidence" from a mathematical, third-person perspective (Meyniel et al., 2015). We would then have arrived at a non-egoic-form of self-modelling, i.e. a slightly richer state that has become "reflexive" in that it begins to continuously track itself. Here is how a phenomenological expert from the Tibetan tradition describes EPISTEMICITY:

> In that moment, we don't see any concrete thing whatsoever. There is an immediate knowing that the essence is empty. There is something that cognizes that the mind is empty, and this cognizant quality is indivisible from the emptiness itself. At the moment that this is an actuality, you don't need to do anything more (T. Rinpoche, 2013, p. 89).

PC6 (TRANSPARENCY/OPACITY) demands that like all other phenomenal representations, MPE can vary along a spectrum of opacity and transparency. The possibility of phenomenal opacity of MPE might help us understand "SELF-LUMINOSITY" or PC3, as noted above. If MPE is fully transparent, the prediction is that there will be an unnoticed absence, because the ARAS-model or wakefulness as such are never explicitly experienced as a representation or something that depends on one's mind – it turns into the mind-independent representation of a global epistemic space (another way to say, "a world"). However, people who

systemically cultivate introspective access should begin to notice brief phases of MPE during global state transitions like waking up.

8.4. What is phenomenality? Towards a conceptual interpretation

The beauty and the difficulty of the "minimal model approach" to consciousness (section 2) lies in the obvious fact that under this approach there is no way in which we can avoid or postpone the *big* question: What, in its essence, *is* phenomenality? I am not going to offer a conceptual model here, because I think a lot of empirical research has to come first. Nevertheless, I will at least sketch one possible route here.

8.4.1 Consciousness as a model of an internal epistemic space

Imagine a congenitally blind person that has just turned on the light. The person notices that many other people present in the same room now suddenly report that they "see things", that they even "see" each other, that there is a strong sense of "togetherness", and that they have all sorts of new capacities. Suddenly, the room is filled with excitement and exchange, with words and movements. In order to understand what has just happened, the blind person must form some sort of idea in her mind – an abstract and global idea of "light" and "visibility". She might arrive at an inner image of some subtle, mysterious, and imperceptible quality of "brightness" and "clarity" pervading the room, or even at a mental representation of her own throwing the light switch having opened the window into an entirely *new* room altogether. The human brain, when it activates itself and is subsequently "woken up" by the ARAS-signal might find itself in a similar situation. And MPE, the experience of consciousness *as such*, might be just this: A global and abstract representation of "epistemic light" and "knowability" that the blind brain has created in order to understand and control its new functional state.

On a more abstract level of analysis, we can interestingly describe consciousness as an epistemic variant of the "self-evidencing brain" (Hohwy, 2016, p. 263). This point is a first building block in developing a new answer to the question *What, exactly, is phenomenality*? Whereas the self-evidencing brain maximizes the sensory evidence for an agent's existence (Friston, 2010, p. 2), the *epistemically* self-evidencing brain first generates the hypothesis *that knowledge is possible*, thereby predicting the occurrence of internal states carrying epistemic value in the future. I think that this could be a first conceptual element for a minimal model explanation of consciousness: MPE is an internal model of *mere epistemic capacity*. The ARAS-model generates predictions about the mean of arousal; thereby creating evidence for the Bayesian "belief" that successful prediction error minimization can be *expected*. However, why should all of this lead to phenomenal experience?

We can now connect this point with the idea that one important marker of conscious processing is "temporal thickness" (Friston, 2018; Limanowski & Friston, 2018, p. 5, 2018). For an organism to have a running ARAS-model is to *continuously* predict a future state of tonic alertness, i.e., to embody a model of the proximate future which includes its own potential to successfully create new epistemic states. As a generative model, the ARAS-model has a *temporal span*, and it simultaneously creates counterfactual depth, because it represents possible future states of knowledge (Friston, 2018, 7, n. 10). It creates functional integration

over time, because the content of the first prediction (of knowledge) causally helps to bring about the physical carrier of the next prediction (of knowledge). This, then, might be the second element in our answer to the question what ultimately makes the ARAS-model a *conscious* state: If we apply the idea of temporal thickness to the special case of epistemic self-evidencing introduced above, then we get a new proposal for what "phenomenality" really is: It is a temporally thick, counterfactually rich representation of the possibility of knowledge. Its concrete phenomenal character (the specific, *sui generis* quality of WAKEFULNESS) is partly determined by the low-level signal itself, while on the level of representational content it is a virtual model of tonic alertness by which the brain "predicts itself into existence as a carrier of epistemic states". Consciousness emerges as the subjective confidence related to a given system's model of its own epistemic capacity.

According to this new model, consciousness – *conscientia* - is not simply "knowing that one knows" (metarepresentation), but rather the probabilistic belief that one *will* know. However, this belief is not some mysterious abstract object, a mere expectation, but a dynamical process which creates a functional link over time (integration). It is internally presented in a special, temporally deep way, a way that continuously connects the present with the organism's proximate future.

Depending on our theoretical interests, what has here been provisionally termed the "ARAS-model" can later be described on different levels of analysis. Let me give three examples.

- If something like the ARAS-model really exists, it can be described as an internal model of an epistemic space (representing the mere capacity or *possibility* of entering knowledge-states, for example via different forms of prediction error minimization like perception, attention, or active inference).
- Second, we might also analyse it as a continuous, temporally thick, and probabilistic representation of *tonic alertness* - if you will, as a virtual model of the very functional process of auto-activation itself, an internal model of the process by which the brain "wakes itself up".
- A third new perspective is that, arguably, it also is the deepest layer of phenomenal embodiment. Because the ARAS-signal is a dynamic body part, the phenomenology of "pure" consciousness quite obviously presents us with an interesting example of bodily self-representation – however, without an instantiation of personal-level self-consciousness or any form of subject-object dichotomy. Perhaps we could even describe the temporally extended and introspectively accessed model of the ARAS-signal as a "subsymbolic token-reflexive", as a body part that continuously represents itself, a neural process that has begun to point to itself. A third, and even more abstract possibility of describing "pure" consciousness therefore is as *a non-egoic form of self-modelling*.

We can now see that the representational space opened by the activated cortex is not mysteriously "aware of itself": The ARAS-model theory of conscious experience is not a reflexive theory because it does not assume that MPE as such is a representation of *itself*, as in countless "awareness-of-awareness" theories. Nor is it a traditional metarepresentational

theory, because prototypical MPE is not conceptually framed as a form of second-order intentional content. Rather, MPE is something slightly different. It is a rather abstract, first-order representation of representational *capacity*; a model of an aperspectival and empty *space* of knowledge. This representation should be highly invariant under counterfactual manipulation, and therefore be typically unnoticed as such.

But again, if it is - what makes all of this *conscious*? If the current proposal points into the right direction, then consciousness is a specific type of non-egoic self-modelling, it appears when a predictive model of an unstructured epistemic space has been embedded into this very space *itself*, beginning to provide internal evidence for the possibility of knowledge.

If MPE can be fruitfully described as a model of such an indivisible epistemic space, having no centre and no boundaries, then the general philosophical picture could be that under standard conditions consciousness is a *transparent* model of an epistemic space. Under standard conditions this model is invisible to the subject of experience, and as the same is true most of the concrete content that later gets superimposed it simply appears as a world that is directly given, which has been known in the past and as a world in which there is a certain probability of new epistemic states occurring in the future. Through contemplative practice, this model can sometimes be made phenomenally opaque, which means that the model is experienced *as such*.

8.5. Steps toward de-idealization

Please recall what I said in the first paragraph of this discussion paper: Whenever I have used terms like "ARAS", "ARAS-signal", or "ARAS-model" these terms should be read as *placeholders*, with each of them specifying a whole range of potential empirical research programs. For example, "ARAS" is the placeholder for the hidden cause of the ARAS-signal, which the ARAS-model predicts and sometimes controls. Of course, it is quite unlikely that what we will ultimately isolate as the neuroanatomical correlate is really what – today - we still call the "ARAS", which is a rather old term and quite possibly not fine-grained enough as a description of the relevant ensemble of brain stem structures. Our views of what exactly the minimally sufficient cause of the relevant signal is at the level of neuroanatomy or connectivity patterns will of course change in the future (see for example Edlow et al., 2012). Every placeholder opens the door for a range of empirical research programs, and as our knowledge of actual mechanisms and causal pathways becomes more fine-grained, our minimal model of "consciousness as such" or MPE will gradually become *de-idealized* and turned into an interdisciplinary research program.

For example, considering the proposed (and highly speculative) switch from model-free to model-based control of the "ARAS-signal", our fine-grained knowledge of the microcircuits actually inducing a global state of sustained cortical arousal can be expected to become more and more detailed (Herrera et al., 2016). Accordingly, our scientific understanding of the fine-grained control mechanisms will change. The same point holds for the functional neuroanatomy relevant to intrinsic and phasic alertness or to the transition from what I have called the functional property of mere "tonic alertness" to the phenomenal property of "wakefulness" (see for example Sturm & Willmes, 2001), or the state-specific architecture of

specific subnetworks (Halassa et al., 2014), or the first candidates for specific networks actually subserving sustained alertness (Coste & Kleinschmidt, 2016; Sadaghiani & D'Esposito, 2015).

On the level of computational modelling, the same "placeholder principle" is true of the simple predictive processing model of MPE I have offered above (see figure 4). As it is part of an idealized minimal model, the computational architecture I have proposed is maximally parsimonious. It has no hierarchical depth, because it assumes only one single level of Bayesian optimization. But alternatives are possible, and nature often does not choose the simplest or most parsimonious option. For example, there could be a *precision model* of the ARAS-signal that can be constantly sustained, independently of whether the ARAS-signal is strong (as in waking states) or weak (as in deep sleep), and of course we could think of system states in which an ARAS-model and a precision model interact and form part of a larger control architecture.[13]

As the main function of this as a discussion paper was to provide first sketch of the ARAS-model of conscious experience and MPE, I will refrain from pointing to any further routes to de-idealization. I hope to provide more detailed answers to the many remaining conceptual and methodological issues, additional phenomenological case studies, and references to historical sources in later versions.

Acknowledgements

Version 8
Monday, 29 April 2019

Contact
metzinger@uni-mainz.de

---

[13] I owe this possibility to Wanja Wiese, personal communication.

References

Albahari, M. (2009). Witness-consciousness: Its definition, Appearance and reality. *Journal of Consciousness Studies*, *16*(1), 62–84.

Alexander, Charles. (1988). A conceptual and phenomenological analysis of pure consciousness during sleep. *Lucidity Letter*, *7*(2).

Apel, K.-O. (1998). *Auseinandersetzungen in Erprobung des transzendentalpragmatischen Ansatzes*: Suhrkamp Frankfurt am Main.

Ataria, Y., Dor-Ziderman, Y., & Berkovich-Ohana, A. (2015). How does it feel to lack a sense of boundaries? A case study of a long-term mindfulness meditator. *Consciousness and Cognition*, *37*, 133–147.

Bachmann, J. (2014). *Accounting for Pure Consciousness: An Examination of the Ability of the Representationalist Approach to Phenomenal Consciousness to Account for Pure Consciousness Experiences*. Edmonton: University of Alberta.

Bachmann, T., & Hudetz, A. G. (2014). It is time to combine the two main traditions in the research on the neural correlates of consciousness: C = L × D. *Frontiers in Psychology*, *5*, 940. https://doi.org/10.3389/fpsyg.2014.00940

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews. Neuroscience*, *16*(7), 419–429. https://doi.org/10.1038/nrn3950

Batterman, R. W. (2002). Asymptotics and the role of minimal models. *The British Journal for the Philosophy of Science*, *53*(1), 21–38.

Batterman, R. W., & Rice, Collin C. (2014). Minimal model explanations. *Philosophy of Science*, *81*(3), 349–376.

Bayne, T., Hohwy, J., & Owen, A. M. (2016). Are There Levels of Consciousness? *Trends in Cognitive Sciences*, *20*(6), 405–413. https://doi.org/10.1016/j.tics.2016.03.009

Blanke, O., & Metzinger, T. (2009). Full-body illusions and minimal phenomenal selfhood. *Trends in Cognitive Sciences*, *13*(1), 7–13.

Brentano, F. (1973). Psychology from an empirical standpoint (AC Rancurello, DB Terrell, & LL McAlister, Trans.). *New York: Humanities.*

Brentano, F. (2012). *Descriptive psychology*: Routledge.

Britton, W. B., Lindahl, J. R., Cahn, B. R., Davis, Jake H., & Goldman, R. E. (2014). Awakening is not a metaphor: The effects of Buddhist meditation practices on basic wakefulness. *Annals of the New York Academy of Sciences*, *1307*, 64–81. https://doi.org/10.1111/nyas.12279

Chatterjee, T. (1982). The concept of Sāksin. *Journal of Indian Philosophy*, *10*(4), 339–356.

Coste, C. P., & Kleinschmidt, A. (2016). Cingulo-opercular network activity maintains alertness. *NeuroImage*, *128*, 264–272. https://doi.org/10.1016/j.neuroimage.2016.01.026

Dentico, D., Ferrarelli, F., Riedner, B. A., Smith, R., Zennig, C., Lutz, A., . . . Davidson, R. J. (2016). Short Meditation Trainings Enhance Non-REM Sleep Low-Frequency Oscillations. *PLoS One*, *11*(2), e0148961. https://doi.org/10.1371/journal.pone.0148961

Dresler, M., & Metzinger, T. (in preparation). Minimal Phenomenal Experience (Part 3): Lucid Dreamles Sleep.

Dretske, F. (1986). Misrepresentation. In R. Bogdan (Ed.), *Belief: Form, content, and function.* Oxford: Clarendon Press.

Dretske, Fred. (1983). Précis of Knowledge and the Flow of Information. *Behavioral and Brain Sciences*, *6*(1), 55–63.

Edlow, B. L., Takahashi, E., Wu, O., Benner, T., Dai, G., Bu, L., . . . Kinney, H. C. (2012). Neuroanatomic connectivity of the human ascending arousal system critical to consciousness and its disorders. *Journal of Neuropathology & Experimental Neurology*, *71*(6), 531–546.

Farrow, J. T., & Hebert, R. J. (1982). Breath Suspension During the Transcendental Meditation Technique. *Psychosomatic Medicine*, *44*(2), 133–153. Retrieved from https://journals.lww.com/psychosomaticmedicine/Fulltext/1982/05000/Breath_Suspension_During_the_Transcendental.1.aspx

Fasching, W. (2011). 'I am of the nature of Seeing': Phenomenological Reflections on the Indian Notion of Witness-Consciousness. In M. Siderits, E. Thompson, & D. Zahavi (Eds.), *Self, no self?: Perspectives from analytical, phenomenological, and Indian traditions.* OUP Oxford.

Ferrarelli, F., Smith, R., Dentico, D., Riedner, B. A., Zennig, C., Benca, R. M., . . . Tononi, G. (2013). Experienced mindfulness meditators exhibit higher parietal-occipital EEG gamma activity during NREM sleep. *PLoS One*, *8*(8), e73417.

Finnigan, B. (2018). Is consciousness reflexively self-aware? A Buddhist analysis. *Ratio*, *31*(4), 389–401. https://doi.org/10.1111/rati.12200

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906.

Forman, R. K. C. (1993). Mystical knowledge: Knowledge by identity. *Journal of the American Academy of Religion*, 705–738.

Forman, R. K. C. (1998). What does mysticism have to teach us about consciousness? *Journal of Consciousness Studies*, *5*(2), 185–201.

Fort, A. O. (1980). The Concept of Suṣupta in Advaita Vedānta. *Annals of the Bhandarkar Oriental Research Institute*, *61*(1/4), 221–228.

Fort, A. O. (1984). The concept of Sāksin in advaita Vedānta. *Journal of Indian Philosophy*, *12*(3), 277–290.

Fox, K. C. R., & Christoff, K. (Eds.). (2018). *The Oxford Handbook of Spontaneous Thought: Mind-wandering, creativity, and dreaming*: Oxford University Press.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127.

Friston, K. (2018). Am I Self-Conscious? (Or Does Self-Organization Entail Self-Consciousness?). *Frontiers in Psychology*, *9*, 579. https://doi.org/10.3389/fpsyg.2018.00579

Frith, C. D., & Metzinger, Thomas. (2016). How the Stab of Conscience Made Us Really Conscious. In A. K. Engel, K. J. Friston, & D. Kragic (Eds.), *The pragmatic turn: Toward action-oriented views in cognitive science* (p. 193). MIT Press.

Gamma, A., & Metzinger, T. (forthcoming). Minimal Phenomenal Experience (Part 2): Psychometric dimensions of "pure consciousness" experiences.

Garrison, K., Santoyo, J., Davis, Jake, Thornhill, T., Kerr, C., & Brewer, Judson. (2013). Effortless awareness: Using real time neurofeedback to investigate correlates of posterior cingulate cortex activity in meditators' self-report. *Frontiers in Human Neuroscience*, *7*, 440.

Gennaro, R. (2008). Are there pure conscious events? In C. Chakrabarti & G. Haist (Eds.), *Revisiting Mysticism* (pp. 100–120).

Giacino, J. T., Ashwal, S., Childs, N., Cranford, R., Jennett, B., Katz, D. I., . . . Zafonte, R. D. (2002). The minimally conscious state definition and diagnostic criteria. *Neurology*, *58*(3), 349–353.

Gupta, B. (1998). *The Disinterested Witness: A Fragment of Advaita Vedānta Phenomenology*: Northwestern University Press.

Halassa, M. M., Chen, Z., Wimmer, R. D., Brunetti, P. M., Zhao, S., Zikopoulos, B., . . . Wilson, M. A. (2014). State-dependent architecture of thalamic reticular subnetworks. *Cell*, *158*(4), 808–821. https://doi.org/10.1016/j.cell.2014.06.025

Harding, D. E. (2000). *Face to No-face: Rediscovering Our Original Nature*: InnerDirections Pub.

Hasenkamp, W. (2018). Catching the wandering mind: Meditation as a window into spontaneous thought. In K. C. R. Fox & K. Christoff (Eds.), *The Oxford Handbook of Spontaneous Thought: Mind-wandering, creativity, and dreaming* (pp. 539–551). Oxford University Press.

Hasenkamp, W., Wilson-Mendenhall, C. D., Duncan, E., & Barsalou, L. W. (2012). Mind wandering and attention during focused meditation: a fine-grained temporal analysis of fluctuating cognitive states. *NeuroImage*, *59*(1), 750–760. https://doi.org/10.1016/j.neuroimage.2011.07.008

Hellie, B. (2007). That Which Makes the Sensation of Blue a Mental Fact: Moore on Phenomenal Relationism. *European Journal of Philosophy*, *15*(3), 334–366. https://doi.org/10.1111/j.1468-0378.2007.00274.x

Herrera, C. G., Cadavieco, M. C., Jego, S., Ponomarenko, A., Korotkova, T., & Adamantidis, A. (2016). Hypothalamic feedforward inhibition of thalamocortical network controls arousal and consciousness. *Nature Neuroscience*, *19*(2), 290–298. https://doi.org/10.1038/nn.4209

Hobson, A., & McCarley, R. W. (1977). The brain as a dream state generator: an activation-synthesis hypothesis of the dream process. *Am J Psychiatr*, *134*, 1335–1348.

Hobson, J., Pace-Schott, E. F., & Stickgold, R. (2000). Dreaming and the brain: toward a cognitive neuroscience of conscious states. *Behavioral and Brain Sciences*, *23*(6), 793–842.

Hohwy, J. (2016). The self-evidencing brain. *Noûs*, *50*(2), 259–285.

James, W. (1950). *Principles of Psychology. Vol. I. New York: Henry Holt, 1890; reprint ed*: New York: Dover.

Laureys, S., Boly, M., Moonen, G., & Maquet, P. (2009). Two dimensions of consciousness: Arousal and awareness. In L. R. Squire, N. F. Dronkers, & J. V. Baldo (Eds.), *Encyclopedia of neuroscience* (Vol. 2, pp. 1133–1142). Elsevier.

Laureys, Steven, Celesia, G. G., Cohadon, F., Lavrijsen, J., León-Carrión, J., Sannita, W. G., . . . Dolce, G. (2010). Unresponsive wakefulness syndrome: A new name for the vegetative state or apallic syndrome. *BMC Medicine*, *8*, 68. https://doi.org/10.1186/1741-7015-8-68

Laureys, Steven, Owen, A. M., & Schiff, N. D. (2004). Brain function in coma, vegetative state, and related disorders. *The Lancet Neurology*, *3*(9), 537–546.

Limanowski, J., & Friston, K. (2018). 'seeing the Dark': Grounding Phenomenal Transparency and Opacity in Precision Estimation for Active Inference. *Frontiers in Psychology*, *9*, 643. https://doi.org/10.3389/fpsyg.2018.00643

Lingpa, K. (2014). *The Tibetan Book of the Dead*: Lulu Press, Inc.

Lutz, A., Slagter, H. A., Dunne, J. D., & Davidson, R. J. (2008). Attention regulation and monitoring in meditation. *Trends in Cognitive Sciences*, *12*(4), 163–169.

Lutz, A., Slagter, H. A., Rawlings, N. B., Francis, A. D., Greischar, L. L., & Davidson, R. J. (2009). Mental training enhances attentional stability: Neural and behavioral evidence. *The Journal of Neuroscience : the Official Journal of the Society for Neuroscience*, *29*(42), 13418–13427. https://doi.org/10.1523/JNEUROSCI.1614-09.2009

MacKenzie, M. (2008). Self-Awareness without a Self: Buddhism and the Reflexivity of Awareness. *Asian Philosophy*, *18*(3), 245–266. https://doi.org/10.1080/09552360802440025

Margolis, E., & Laurence, S. (2014). Concepts. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (2014th ed.). Metaphysics Research Lab, Stanford University.

Maruthai, N., Nagendra, R. P., Sasidharan, A., Srikumar, S., Datta, K., Uchida, S., & Kutty, B. M. (2016). Senior Vipassana Meditation practitioners exhibit distinct REM sleep organization from that of novice meditators and healthy controls. *International Review of Psychiatry*, *28*(3), 279–287.

Mason, L., Alexander, Charles N., Travis, Frederick T., Marsh, G., Orme-Johnson, D., Gackenbach, J., . . . Walton, K. G. (1997). Electrophysiological Correlates of Higher States of Consciousness During Sleep in Long-Term: Practitioners of the Transcendental Meditation Program. *Sleep*, *20*(2), 102–110.

Mason, L., & Orme-Johnson, David. (2010). Transcendental consciousness wakes up in dreaming and deep sleep. *Int J Dream Res*, *3*, 28–32.

Melchert, N. (1977). Mystical Experience and Ontological Claims. *Philosophy and Phenomenological Research*, *37*(4), 445–463.

Merrell-Wolff, F. (1973). *Pathways through to space*: New York: Julian Press.

Metcalfe, J., & Shimamura, A. P. (Eds.). (1996). *Metacognition: Knowing about knowing* (1st MIT Press paperback ed.). Cambridge, Mass: MIT Press.

Metzinger, T. (2003a). *Being No One: The Self-model Theory of Subjectivity*: MIT Press.

Metzinger, T. (2015a). M-Autonomy. *Journal of Consciousness Studies*, *22*(11-12), 270–302.

Metzinger, T. (2003b). Phenomenal transparency and cognitive self-reference. *Phenomenology and the Cognitive Sciences*, *2*(4), 353–393.

Metzinger, T. (Ed.). (2000). *Bradford Books. Neural correlates of consciousness: Empirical and conceptual questions*. Cambridge, Mass: MIT Press.

Metzinger, T. (2006). Conscious volition and mental representation: Toward a more fine-grained analysis. In N. Sebanz & W. Prinz (Eds.), *Disorders of volition* (pp. 19–48). Cambridge, Mass: MIT Press.

Metzinger, T. (2013). The myth of cognitive agency: Subpersonal thinking as a cyclically recurring loss of mental autonomy. *Frontiers in Psychology*, *4*, 931.

Metzinger, T. (2014). How does the brain encode epistemic reliability? Perceptual presence, phenomenal transparency, and counterfactual richness. *Cognitive Neuroscience*, *5*(2), 122–124. https://doi.org/10.1080/17588928.2014.905519

Metzinger, T. (2017). The Problem of Mental Action. In T. K. Metzinger & W. Wiese (Eds.), *Philosophy and Predictive Processing.* Frankfurt am Main: MIND Group. https://doi.org/10.15502/9783958573208

Metzinger, T. (2018). Why is mind wandering interesting for philosophers? In K. C. R. Fox & K. Christoff (Eds.), *The Oxford Handbook of Spontaneous Thought: Mind-wandering, Creativity, and Dreaming* (pp. 97–111). Oxford University Press.

Metzinger, T. (2019, submitted). Minimal Phenomenal Experience (Part 1): Minimal Phenomenal Experience (Part 1) Meditation, tonic alertness, and the phenomenology of "pure" consciousness.

Metzinger, Thomas, & Windt, Jennifer Michelle. (2014). Die phänomenale Signatur des Wissens: Experimentelle Philosophie des Geistes mit oder ohne Intuitionen? In T. Grundmann, J. Horvath & J. Kipper (Ed.), *Die Experimentelle Philosophie in der Diskussion* (pp. 279–321). Suhrkamp.

Metzinger, Thomas, & Windt, Jennifer M. (2015). What Does it Mean to Have an Open MIND? In T. K. Metzinger & J. M. Windt (Eds.), *Open MIND.* Frankfurt am Main: MIND Group. https://doi.org/10.15502/9783958571044

Meyniel, F., Sigman, M., & Mainen, Z. F. (2015). Confidence as Bayesian Probability: From Neural Origins to Behavior. *Neuron*, *88*(1), 78–92. https://doi.org/10.1016/j.neuron.2015.09.039

Moore, G. E. (1903). The refutation of idealism. *Mind*, *12*(48), 433–453.

Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, *83*(4), 435–450.

Nagel, T. (1986). The View from Nowhere. *Oxford: Oxford University Press*, *1*, 86.

Namgyal, D. (2006). *Mahamudra - The moonlight: Quintessence of mind and meditation*: Wisdom Publications,u.s.

Nash, J. D., & Newberg, A. (2013). Toward a unifying taxonomy and definition for meditation. *Frontiers in Psychology*, *4*, 806.

Oken, B. S., Salinsky, M. C., & Elsas, S. M. (2006). Vigilance, alertness, or sustained attention: physiological basis and measurement. *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology*, *117*(9), 1885–1901. https://doi.org/10.1016/j.clinph.2006.01.017

Pekala, R. J. (1991). The phenomenology of meditation. In M. A. West (Ed.), *Oxford science publications. The psychology of meditation* (pp. 59–80). Oxford: Clarendon Press. https://doi.org/10.1093/acprof:oso/9780198521945.003.0004

Peters, F. (2013). Theories of Consciousness as Reflexivity. *The Philosophical Forum*, *44*(4), 341–372. https://doi.org/10.1111/phil.12018

Picard, F. (2013). State of belief, subjective certainty and bliss as a product of cortical dysfunction. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *49*(9), 2494–2500. https://doi.org/10.1016/j.cortex.2013.01.006

Posner, M. I. (2008). Measuring alertness. *Annals of the New York Academy of Sciences*, *1129*, 193–199. https://doi.org/10.1196/annals.1417.011

Rgyal-ba-g'yung-drung, B.-s., Gurung, G. S., & Brown, D. (2017). *The Pith instructions for the stages of the practice sessions of the a khrid system of Bon Rdzogs Chen (great completion) meditation* (Second edition). [Occidental, CA]: Bright Alliance.

Rice, Collin, Rohwer, Y., & Ariew, A. (2018). Explanatory schema and the process of model building. *Synthese*, 1–23.

Rinpoche, K., & Namgyal, L. (2011). *The Ninth Karmapa's Ocean of Definitive Meaning*: Snow Lion Publications, Incorporated.

Rinpoche, T. (2000). *As it is, Volume II. Rangjung Yeshe books*. Hong Kong: Rangjung Yeshe Publications.

Rinpoche, T. (2013). *As It Is, Volume I*. Hong Kong: Rangjung Yeshe Publications.

Roediger III, H., & Marsh, E. (2009). False memory. *Scholarpedia*, *4*(8), 3858. https://doi.org/10.4249/scholarpedia.3858

Russell, I. F. (2013). The ability of bispectral index to detect intra-operative wakefulness during total intravenous anaesthesia compared with the isolated forearm technique. *Anaesthesia*, *68*(5), 502–511.

Sadaghiani, S., & D'Esposito, M. (2015). Functional Characterization of the Cingulo-Opercular Network in the Maintenance of Tonic Alertness. *Cerebral Cortex (New York, N.Y. : 1991)*, *25*(9), 2763–2773. https://doi.org/10.1093/cercor/bhu072

Sanders, R. D., Tononi, G., Laureys, Steven, & Sleigh, J. W. (2012). Unresponsiveness≠ unconsciousness. *Anesthesiology: the Journal of the American Society of Anesthesiologists*, *116*(4), 946–959.

Schoenberg, P., Ruf, A., Churchill, J., Brown, D., & Brewer, Judson A. (2018). Mapping complex mind states: EEG neural substrates of meditative unified compassionate awareness. *Consciousness and Cognition*, *57*, 41–53.

Seli, P., Kane, M. J., Metzinger, Thomas, Smallwood, J., Schacter, D. L., Maillet, D., . . . Smilek, D. (2018b). The Family-Resemblances Framework for Mind-Wandering Remains Well Clad. *Trends in Cognitive Sciences*, *22*(11), 959–961. https://doi.org/10.1016/j.tics.2018.07.007

Seli, P., Kane, M. J., Smallwood, J., Schacter, D. L., Maillet, D., Schooler, J. W., & Smilek, D. (2018a). Mind-wandering as a natural kind: A family-resemblances view. *Trends in Cognitive Sciences*, *22*(6), 479–490.

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, *17*(11), 565–573. https://doi.org/10.1016/j.tics.2013.09.007

Seth, A. K. (2015). The Cybernetic Bayesian Brain. In T. K. Metzinger & J. M. Windt (Eds.), *Open MIND*. Frankfurt am Main: MIND Group. https://doi.org/10.15502/9783958570108

Severeide, C. J. (1990). Physiological and phenomenological aspects of transcendental meditation. *Scientific Research on Maharishi S Transcendental Meditation and TM-Sidhi/'rogram, Maharishi European Research University Publication*. (3121), l556-84.

Shear, J. (2007). Eastern methods for investigating mind and consciousness. In S. Schneider & M. Velmans (Eds.), *The Blackwell companion to consciousness* (pp. 697–710). John Wiley & Sons.

Siegel, S. (2007). How can we discover the contents of experience? *The Southern Journal of Philosophy*, *45*(S1), 127–142.

Stace, W. T. (1960). *Mysticism and philosophy*. London: Macmillan.

Steriade, M. (1996a). Arousal--Revisiting the Reticular Activating System. *Science*, *272*(5259), 225-0. https://doi.org/10.1126/science.272.5259.225

Steriade, M. (1996b). Awakening the brain. *Nature*, *383*(6595), 24.

Steriade, Mircea, & McCarley, R. W. (2013). *Brainstem control of wakefulness and sleep*: Springer Science & Business Media.

Sturm, W., & Willmes, K. (2001). On the functional neuroanatomy of intrinsic and phasic alertness. *NeuroImage*, *14*(1 Pt 2), S76-84. https://doi.org/10.1006/nimg.2001.0839

Sullivan, P. R. (1995). Contentless consciousness and information-processing theories of mind. *Philosophy, Psychiatry, & Psychology*, *2*(1), 51–59.

Thompson, E. (2015b). *Waking, dreaming, being: Self and consciousness in neuroscience, meditation, and philosophy*. New York and Chichester: Columbia University Press.

Thompson, E. (2015). Dreamless Sleep, the Embodied Mind, and Consciousness. In T. K. Metzinger & J. M. Windt (Eds.), *Open MIND.* Frankfurt am Main: MIND Group. https://doi.org/10.15502/9783958570351

Travis, Fred, & Shear, J. (2010). Focused attention, open monitoring and automatic self-transcending: Categories to organize meditations from Vedic, Buddhist and Chinese traditions. *Consciousness and Cognition*, *19*(4), 1110–1118. https://doi.org/10.1016/j.concog.2010.01.007

Travis, Frederick, & Pearson, C. (2000). Pure consciousness: Distinct phenomenological and physiological correlates of "consciousness itself". *International Journal of Neuroscience*, *100*(1-4), 77–89.

Wainwright, W. J. (1981). *Mysticism: A study of its nature, cognitive value and moral implications*. Madison: University of Wisconsin Press.

Ward, A. F., & Wegner, D. M. (2013). Mind-blanking: When the mind goes away. *Frontiers in Psychology*, *4*, 650. https://doi.org/10.3389/fpsyg.2013.00650

Weinbach, N., & Henik, A. (2012). The relationship between alertness and executive control. *Journal of Experimental Psychology. Human Perception and Performance*, *38*(6), 1530–1540. https://doi.org/10.1037/a0027875

Weisberg, M. (2012). *Simulation and similarity: Using models to understand the world*: Oxford University Press.

Wiese, W. (2018). *Experienced Wholeness. Integrating Insights from Gestalt Theory, Cognitive Neuroscience, and Predictive Processing*. Cambridge, MA: MIT Press.

Wiese, W., & Metzinger, Thomas K. (2017). Vanilla PP for Philosophers: A Primer on Predictive Processing. In T. K. Metzinger & W. Wiese (Eds.), *Philosophy and Predictive Processing.* Frankfurt am Main: MIND Group. https://doi.org/10.15502/9783958573024

Williams, P. (1998). *The reflexive nature of awareness: A Tibetan Madhyamaka defence. Curzon critical studies in Buddhism: [1]*. Richmond: Curzon.

Williford, K., Bennequin, D., Friston, K., & Rudrauf, D. (2018). The Projective Consciousness Model and Phenomenal Selfhood. *Frontiers in Psychology*, *9*, 732. https://doi.org/10.3389/fpsyg.2018.02571

Windt, Jennifer. (2015). *Dreaming: A Conceptual Framework for Philosophy of Mind and Empirical Research*. Cambridge, MA: MIT Press.

Windt, Jennifer M. (2015). Just in Time—Dreamless Sleep Experience as Pure Subjective Temporality. In T. K. Metzinger & J. M. Windt (Eds.), *Open MIND.* Frankfurt am Main: MIND Group. https://doi.org/10.15502/9783958571174

Windt, Jennifer M., Nielsen, T., & Thompson, E. (2016). Does consciousness disappear in dreamless sleep? *Trends in Cognitive Sciences*, *20*(12), 871–882.

Wislowska, M., del Giudice, R., & Schabus, M. (2014). Searching for Neuronal Correlates of Consciousness-What Do We Learn from Coma and Disorders of Consciousness. *Functional Neurology, Rehabilitation, and Ergonomics*, *4*(2/3), 135.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology*, *18*(5), 459–482.