# A mathematical model of embodied consciousness

David Rudrauf[1], Daniel Bennequin[2], Isabela Granic[3], Gregory Landini[4], Karl Friston[5], Kenneth Williford[6]

**1** CISA/FAPSE/Campus Biotech, University of Geneva, Geneva, Switzerland
**2** Department of Mathematics, University of Paris 7, Paris, France
**3** Department of Behavioral Science, Radboud University, Nijmegen, Netherlands
**4** Department of Philosophy, University of Iowa, Iowa City, USA
**5** Wellcome Trust Center for Neuroimaging, University College London, London, UK
**6** Department of Philosophy, University of Texas, Arlington, USA

* david.rudrauf@unige.ch

## Abstract

We introduce a mathematical model of embodied consciousness, the Projective Consciousness Model (PCM), which is based on the hypothesis that the spatial field of consciousness (FoC) is structured by a projective geometry and under the control of a process of active inference. The FoC in the PCM combines multisensory evidence with prior beliefs in memory, for the choices of projective points of view, that respond to a combination of spatial and affective priors. Free energy minimization drives intentionality, perspective taking and the switch between perception, imagination and action in the FoC. Changes of projective frames govern how expectations are transformed by consciousness. Our model suggests that changes of brain states from unconscious to conscious reflects the action of projective transformations. In the PCM, consciousness functions as an algorithm for the maximization of resilience, using projective perspective taking and imagination in order to escape local minima of free energy. The PCM can explain a variety of psychological phenomena: the manifestation of subjective experience with its characteristic spatial phenomenology, the distinctions and integral relationships between perception, imagination, and action, the role of affective processes in intentionality, but also perceptual phenomena such as the dynamics of bistable figures and body swap illusions in virtual reality. It relates phenomenology to function, showing its computational and adaptive advantages. It suggests specific neurophenomenological hypotheses about the brain, guidelines for designing artificial systems, and formal principles for psychology.

## 1 Introduction

### 1.1 Basic rationale and contextualization

We introduce a mathematical model of embodied consciousness, the Projective Consciousness Model (PCM), that identifies core aspects of the phenomenology of consciousness with specific computational mechanisms. The model integrates active inference and projective geometry to define a Field of Consciousness (Foc). The FoC operates as a mechanism of arbitration based on the anticipation of the consequences of action. Its architecture embeds mechanisms of cognitive and affective appraisal and reappraisal that exploit both local perceptual inference and non-local projective imagination in order to orient attention and motivate action. The overall algorithm yields a model that optimizes resilience to stressors and adverse conditions. The optimization of resilience has been hypothesized to be a core function for embodied consciousness [1]. The PCM

explains a broad range of perceptual, cognitive, and affective phenomena and leads to straightforward neurocomputational hypotheses with a broad application potential.

The PCM is based on an interdisciplinary synthesis at the crossroads of multiple traditions and concepts: Phenomenology [2–5], Francisco Varela's neurophenomenology program and reflection on embodiment and autonomy [6–9], Karl Friston's theory of active inference based on free energy minimization as a model of embodied autonomous agents [10], Alain Berthoz's emphasis on the sense of space as a multisensory, supramodal phenomenon [11]; projective geometry, from the geometry of art and architecture during the Renaissance period to the theory of abstract cameras and the latest advances in computer graphics [12, 13], virtual reality, as it brings new insights about self-consciousness [14], but also past and current models of consciousness [15, 15–25], neuropsychology, computational neuroscience [1, 26, 27] and the affective sciences [28–31].

Our aim is not to try to solve the so-called hard problem [32], but to focus on the structural phenomenology of experience and its representation in an overall computational framework capable of explaining complex behaviors, decision-making and action, based on internal parameters. We aim at establishing based on general mathematical and computational principles, how we, as conscious, embodied systems, perceive, imagine, and act in the 3-dimensional spatial world, from a first person perspective, through a structural coupling with our environment, in order to cope with physical and social situations and maintain our autonomy by optimizing resilience [1, 6, 8, 10]. Along with other models, we conceive of the core of consciousness as an integrative global workspace [15, 15, 16, 33], selected through evolution for adaptive representation and control in an embodied context. We assume that consciousness participates in the orientation and directedness of cognition and action towards concrete and abstract objects in accordance with preferences; this can be expressed by the claim that consciousness is structured by an "intentional function" or intentionality [34]. More generally, conscious agents have to solve the problem of covertly or overtly (by explicit bodily actions) bringing into focus various parts of their surroundings for information integration via sensory evidence and simulation in order to navigate the world. Such intentionality can manifest at a psychological level, for example, as the so-called "attentional beam" [15]. Our framework is also compatible with the general hypothesis that consciousness has evolved as a mechanism that optimizes information integration [17, 35], even though we do not formally relate our model to this framework here. Furthermore, we do not attempt to elucidate the relationship between so-called "access consciousness" and "phenomenal consciousness" (see [36]) in the context of our model. Our view is that (1) phenomenology (or qualitative character, as it is also sometimes called, e.g., of a tomato on a table, presenting the features of its exposed surface to perception) normally correlates with (2) related intentional content (e.g., a tomato's-being-on-the-table) and (3) accessibility (in the global workspace) for further actions (e.g., I can report what I see and how it looks from my position, decide to pick up the tomato depending on my current narrative). These components may be dissociable to some extent, but one can think of our model as framing them as integrated in the overall structure of human conscious experience.

We place an emphasis on the embodied nature of consciousness because we see embodiment as a strong determinant of its overall structure and dynamics [6]. However, our model does not reduce consciousness to embodiment in a strong sense, as the workspace relies on multiple layers of abstraction to represent and control the body, and the model is compatible with multiple forms of embodiment. Nevertheless, embodied cognition constitutes a framework within which a theory of consciousness can be derived in an ecological way. Autonomous embodied cognitive systems can be described as situated agents that operate from a certain location in a specific environment under particular systemic constraints [6]. They can further be described as performing perceptual and active inference in order to accumulate and deploy adaptive knowledge about themselves and the world through their history of interactions with it [6, 10]. In the framework of active inference, they carry out cycles of perception, prediction, and action according to prior beliefs about the self and the world held in their memory. These priors are constantly being updated and optimized, increasing the predictive power of internal models of physical and social exchanges with the world (e.g., energy efficiency, group acceptance). The internal models are used to appraise sensory

evidence in relation to expectations and preferences and to govern action, including behaviors of approach and avoidance [1, 31, 37, 38]. The optimization of this overall process of appraisal will increase the agent's likelihood of remaining within its domain of viability (homeostasis) and well-being, and help it cope with adverse events by maximizing its resilience [1, 39].

One of the pivotal consideration for the derivation of the PCM is that for such embodied systems in general, there is an evolutionary advantage of developing an integrative cognition of space in order to represent, simulate, appraise and control spatially distributed information and the consequences of actions. Geometries and spaces are important components of the distribution of information the agents' brains need to determine in order to model relative changes in the internal and external environments. This need is so engrained that sensory organs have evolved so as to embed implicit geometrical assumptions (e.g., the vestibular and visual systems), and spatial cognition has undergone selective pressure throughout evolution to incorporate models of complex spatial relationships between agents, in the context of the assessment of personally relevant outcomes (e.g., in prey and predator dynamics and luring strategies) [40]. The need for making inferences about space strongly suggests the existence of internal predictive models operating over sensory experience and prior beliefs in memory, through which the world is represented and interpreted from a situated perspective.

Predictive models about spatial information can be expected across the hierarchy of neurocognitive processing, including non-conscious levels. But we hypothesize that, at the tip of the iceberg, consciousness relies on an integrative predictive model attributing a global 3-dimensional spatial geometry to multimodal sensory information and memory traces as they access the conscious workspace. We also hypothesize that this model serves as a support for the mapping and appraisal of preferences and personally relevant information. We expect that such a spatial model offers a parsimonious, minimally complex mechanism, which is consistent with basic principles of approximate Bayesian inference (and free energy minimization) but also with Occam's Razor, which governs (statistical or generative) models that optimally generalize to new data [10]. We hypothesize that such global internal model of space operates in a supramodal manner by dynamically integrating and unifying information from multiple sensory modalities through affective arbitration [1].

Of course, there is much more to consciousness than spatiality and its valuation. In a normal human individual, consciousness extends to incorporate differentiated and sophisticated levels of representation [28]: from perception to language and mathematical thought. And lived spatiality must be thought of as integrated with the flow of lived time [7]. But, to begin formulating a mathematical model of consciousness, it is sound from a phenomenological, functional, and practical standpoint to consider core consciousness [28] from the perspective of embodied spatial cognition in the context of active inference, which incorporates complex cognitive dynamics.

Our approach directly connects to a fast growing literature, relating the free energy principle [10], to perceptual inference, active inference and (embodied) conscious experience. The framework of free energy driven perceptual and active inference has been used to shed light on conscious processing, including the structure and unity of consciousness, aspects of the first person perspective, self-models, embodiment, interoception, body illusions, and bistability [20–25]. The PCM advances previous formulations of active inference (in relation to consciousness) by featuring a mathematical theory, based on projective geometry, of the form, structure and dynamics of spatial, perceptual, and imaginary experience. In particular, it accommodates counterfactual or "as if" multi-perspectival transformations of the first-person perspective in conscious experience. It specifies a core mechanism unifying perception, imagination and action for the optimization of the appraisal of action outcomes, and thus maximizes the overall resilience of an embodied cognitive system imbued with it. It offers an explicit, formal, computable and integrative account of embodied consciousness, which links components of the affective sciences, and offers a basis for testing hypotheses about psychological mechanisms quantitatively.

Our argument follows three basic steps.

First, we assume that conscious processing entails a process of inference, in particular, active

inference under the free energy principle. It relates the conscious intentional experience of the world with the products of typically unconscious inference based on sensory evidence.

The second assumption we make follows from the first, namely, that to make inferences or predictions any sentient (or conscious) creature must possess a generative model of its sensorium. In other words, to engage in active inference it is necessary to generate predictions about what would happen if I did such-and-such. This is necessary for testing internal hypotheses against sensory evidence and updating one's beliefs (or behavior) accordingly. This fits comfortably with the Bayesian brain hypothesis [41–43] and more fundamental arguments such as the Good Regulator Theorem (from early formulations of self-organization) [20, 44].

Finally, and most importantly, we posit projective geometry as a necessary form of the generative model that must be deployed by any conscious organism that navigates a three-dimensional world. Put simply, to produce generalizable and veridical predictions, the internal models used by our brain must recapitulate the formal geometry of the world we are trying to make predictions about (from situated standpoints). In this setting, the Projective Consciousness Model (PCM) regards projective geometry (in three dimensions) as the optimal geometry for situated inferences to the best explanation for sensory impressions.

In what follows, we will see that the ability to operate on or within a projective geometry carries with it all the necessary equipment for the sort of active inference characteristic of beings like ourselves, in particular, the ability to entertain fictive outcomes and multiple perspectives on the composition and dynamics of objects causing (or that could cause) sensory input. We focus here on the projective aspects of generative models and the pre-eminent role of projective geometry (as opposed to active inference and free energy per se). We suppose that such a geometry must (at some level) be entailed by the generative models used by the brain (see section on neurocomputational hypotheses).

## 1.2 Phenomenological postulates: a mode of access with a characteristic spatial phenomenology based on perspective taking and driven by the mapping of personally relevant information

We specify how spatiality is consciously experienced, including one's own body and oneself as observer [14]. Our goal is to infer from this analysis, moving from what is represented to what carries out the representing, key computational properties of consciousness that can account for the generation of its characteristic spatial phenomenology. This will later allow us to make hypotheses about its neural implementation.

To ground the model, we use the following well-motivated postulates, which distill centuries of discussions about the structure of subjectivity and consciousness. They feature an integrative subjective experience with a characteristic spatial phenomenology that is accessed from a specific perspective (a "point of view"). They were identified prior to any particular mathematical model from a synthesis of psychology, cognitive neuroscience, phenomenology, and philosophy of mind. Though not uncontroversial, we accept them here as a plausible starting point from which to draw out important consequences. (A similar method can be found in [35].)

### 1.2.1 A 3-dimensional lived space

Investigations into the phenomenology of consciousness have emphasized consciousness as grounded in a subjective standpoint and a space of sorts, using the basic schematics of a subject-object opposition and a relation between a center and a periphery [2, 3, 7, 45]. Spatiality is a pervasive aspect of subjective experience. When one closes one's eyes in a silent room, under normal conditions, one experiences a "lived space", to adopt a term of art commonly used in the Phenomenological tradition, the sense of a 3-dimensional surrounding space that is invariant even as the state of the outside world constantly changes. We find such a space in both waking consciousness and in dream. It seems central to conscious experience in general, to the way it represents a world in three dimensions, in which we locate ourselves and from which we perceive,

imagine, and plan action [11, 14, 46–50]. In this context, the particular "way" such a space appears is considered as an immanent structure of the subjective conscious experience we "live through", and to which we have phenomenological access; it is not a feature of the world represented [51] (it only secondarily becomes itself an object of representation, something to which our attention can be drawn, just as one does not normally notice one's hand in hammering, though one can). As a first approximation, lived space can be said to feature, by default, a primarily egocentric representation of space, defining a peri-personal and an extra-personal environment in relation to a body, and oriented along directions of perception and action [11, 46].

It manifests as a unified, supramodal and normally coherent experience, integrating and mapping, into a cohesive and continuous global workspace, different portions of space, as sampled by multiple, redundant sense modalities (notably audition and proprioception), and constantly filled with information from memory to complement the partial sensory sampling (e.g., I'm looking at a stage and listening to the music playing, but I know and feel that my friend is sitting next to me). Vision is central to the process, but it is not necessary to it (e.g., blind subjects also have a rich perception and imagination of objects and of others in space). The visual field is an informationally rich and sharp representation of a portion of the ambient space, which is integrated into the broader, more encompassing conscious experience of a global spatial field or spatiality, with a unified frame of reference and a representation of the body [11, 48, 52]. Thus all sense modalities, vision, audition, proprioception, vestibular processing, olfaction, touch, but also interoception are expected to contribute to its ongoing construction in real-time, and to populate its content adaptively [11, 52].

As suggested above, lived space ought not be considered a direct representation of the intrinsic spatial structure of the world, and can be conceived as an adaptive simulation, "user interface", projection or controlled "hallucination" that is actively constructed by the brain in real-time [10, 48, 53]. Experientially, outside of pathological conditions, it engenders an illusion of direct realism, (i.e., of being conscious of the world itself directly) [54], which is probably a necessary condition for effectively representing our real environment and acting within it. But dreams, imagination, hallucinations, out of body experiences (OBEs), sensory substitution, vertigo, and various neurological syndromes, provide compelling evidence that our lived, moment-by-moment representation and subjective experience of ambient physical space is the construction of a virtual world. In fact, the normal, robust, ongoing conscious experience of co-localization in space of one's subjective point of view and of one's own body, is contingent upon a process of multisensory inference, and reflects a dynamical balance in sensorimotor calibration [14, 55]. Accordingly, when we use terms like 'embodied', this should not be taken to commit us to direct realist, externalist, or "extended-mind" theories of conscious perception (see, for example [56]). In our view, consciousness is a computational process performed entirely within one's brain; even though, clearly, it bears causal connections to processes that extend well beyond the brain and, by intentionality, it represents objects, events, and states of affairs that are certainly not contained within the brain).

### 1.2.2 An elusive standpoint and a space with a representation of the body at the center

In default egocentric spatial consciousness, the center of lived space seems to coincide with an implicit vantage point, somewhat near or intersecting with the perceived location of the head. Phenomenologically, the vantage point has an ill-defined and elusive localization in the space, even though it seems to play a central, if usually tacit role, like that of an origin, in the organization of the space: one can locate objects in space, but one cannot precisely locate the vantage point from which the representation seems to emerge as a simple location in 3-dimensional space [5, 57]. It seems to behave like a virtual, internal 3-dimensional camera, which is reminiscent of the long-debated notion of an "inner eye", the basis of some classical aporias of the theory of consciousness, (e.g., the homuncular regress problem, the problem of the location of the subject, and issues surrounding the "Cartesian Theater") [5, 58–60].

Philosophers have debated over centuries about whether one could find a "subject" in consciousness [61–63]. Such a subject would, among other things, provide an origin and foundation for its lived space. The debate has oscillated between affirming and negating its presence [5, 64]. If it exists, it remains phenomenologically elusive and cannot be accessed like normal objects. However, the intuition is pervasive and difficult to eliminate, and we wish to investigate whether we can make sense of it with a mathematical model, bracketing ontological and epistemological concerns.

Lived space is also characterized by the pervasive presence of a 3-dimensional representation of the body, which monitors the state and controls the action of the actual body. The existence of such a lived representation of the body is supported by evidence from perturbations of proprioception in various experimental and clinical conditions [14], by the pervasive presence of phantom limbs in amputated patients (which the patients can control to a certain extent as they would control the missing limb [65]), and by ordinary mental imagery.

In contrast to most contents of consciousness, the lived body is normally always present in the conscious field, though, of course, not normally the object of explicit attention or conceptual representation. Its integrity, as a proxy for the integrity of the actual body, is critical for an embodied autonomous agent. As a space imbued with value, it is an anchor point for our efforts at preserving autonomy and well-being. Protecting the actual body in physical space requires an ongoing monitoring and appraisal of one's own bodily position and state in space in relation to past, present, and future situations. The lived body is thus a kind of inferential representation of the real body in physical space, reliant upon memory, anticipation, and ongoing multimodal sensorimotor information. It constitutes a sort of virtual "user interface" for the representation and control of the actual body, just like the rest of the global workspace. It can be thought of as a kind of "virtual body avatar", richly yet selectively informed by multiple sources of information [14]. We control our limbs through anticipated proprioceptive representations of their position in space, and we avoid and approach things based on their anticipated potential personal relevance in connection to their impact on the body.

In normal perception, the lived body manifests as a 3-dimensional phenomenon in a default egocentric mode, that is, from a first-person perspective. The representation is informed by proprioception and interoception and by other sense modalities, like vision, audition, and the vestibular sense, but also olfaction; all of these act as spatial filters. The lived body and the real body are related: their referential are "superimposed" via ad hoc calibration for optimal representation and motor control. This implicit superimposition is dramatically revealed by phantom limb phenomena.

In imagination, or in abnormal perceptual situations such as body-swap illusions or normal effects of mirror reflections [14, 55], the body also manifests as a similar 3-dimensional phenomenon, under various egocentric and allocentric modes, from first-person or third-person perspectives. In spite of their frequent relative separation in lived space during imagination, experimental situations, or the confrontation with reflections, both the implicit vantage point and the representation of the body generally remain identified as one's own; they preserve their identification as belonging to the same agent in consciousness.

Such lived representation of one's own body is present in many forms of subjective experience involving an explicit and reflective form of self-consciousness. It is present in mind-wandering, day-dreaming, and fantasy [66]. It is constantly cast, with more or less emphasis, into the ongoing narratives and scenarios that we unfold as subjects through imagination in order to aid us in appraising, interpreting, and manipulating physical and social situations so as to maximize utility, in a normally highly biased way.

### 1.2.3   A mechanism for spatial intentionality and perspective taking

Lived space necessarily involves spatial intentionality (i.e., the ability to relate to objects as a situated observer and to aim at locations in space) and intentionality as an organizing function is central to consciousness [34]. There is a sense that one's awareness of objects goes out from a

subject toward an object [2]. This is in evidence in spatial selective attention and is described as an "attentional spotlight" in the global workspace model [15] (which invokes the metaphors of a 3-dimensional theater with an implicit observer and a stage). In the space of consciousness, points are not without valence or relation; they are the focus of "beams" of attention or the sources of such "beams". This spatial intentionality is independent of overt actions or orientation (e.g., head orientation and saccades). It can operate covertly or overtly across perception, imagination, action planning and self-representation, and acts as a principle of unification among these processes. It is key to the control of behavior [67]. We should add that in speaking of 'intentionality' here we have in mind the general feature of "object-directedness" that is presupposed by all higher forms of representation. We are suggesting that this "directedness" or "aim" is a fundamentally geometrical notion that, of course, develops, in conceptual thought and language, well beyond its roots in perceptual consciousness.

The dynamics of intentionality and orientation in lived space is controlled by a complex ensemble of parameters, which notably relate to biological constraints, and the ongoing appraisal of the cognitive and affective personal relevance of objects, spaces and situations (e.g., we pay attention to what we fear or desire and where it is to be found) [37]. Affective processes, in connection to memory, have long been conceived as playing a key role in driving intentionality and consciousness [2, 7, 8, 28].

Perspective taking, the act of representing space and its contents from a particular standpoint is central to the overall process of comporting oneself in lived space by representing physical and social relationships across different angles [68]. It implements the spatial expression of intentionality in attention. It is itself driven by the detection of personally relevant information, in connection to memory. Perspective taking is the common mechanism by which the frames of reference of perception, imagination, action-programming, and self-representation are related. In perception, the frame of reference of lived space tends to default to an egocentric stance, with a first-person perspective. In imagination, we can adopt any allocentric vantage point: one can imagine something from somebody else's point of view or an object from a first-person perspective, or oneself being represented from a second-person perspective, or oneself relating to another or an object from a third-person perspective. We also can imagine the points of view of non-human entities, from tiny animals to a "God", and from any distance (e.g., "at arm's length", "with a ten-foot pole", "from 30,000 feet").

This implies that possible other perspectives are, in effect, built into lived space. Evolutionary pressure has been placed on animals to represent themselves from the perspective of another [69] e.g., the prey as seen by the predator and conversely, which has enabled animals to develop luring strategies based on anticipation, and to create and exploit perspectival illusions for mating success [70]. In our model, access to representations of space in consciousness always implies perspective taking. Creatures with no capabilities for third-person point of view perspective taking would have a very limited consciousness, as they would be unable to use anticipatory local and non-local perspective taking to build an integrative representation of their surroundings; thereby escaping their situated first-person perspective..

### 1.2.4   A workspace in perspective with a horizon

In both normal perception and imagination, the experience of spatiality generally follows specific rules of transformation in which the sense of direction (front, back, up, down, left, right) and of relations of incidence are preserved. The alignment of objects along a straight line is preserved in changes of point of view (e.g., when perceiving or imagining roads under different points of view). Likewise, when seeing or imagining an object moving, we normally perceive the incidences of its parts as being preserved; otherwise the object would lose its unity. From an evolutionary standpoint, representing objects in a way that preserves relations of incidence is a critical adaptive feature in a world of inertia, collisions, and encounters. Relations of incidence are personally relevant for embodied systems.

In lived space, objects, points of view and their relations of incidence are represented with the

shape of 3-dimensional projections in perspective, along certain directions, from a certain standpoint; parallel lines converge on an implicit or explicit horizon at infinity. It is through the lens of this inherently perspectival geometry that a world of transcendent 3-dimensional objects in Euclidean space existing in independence of any particular perspectives (whose sizes remain the same irrespective of distance) is consciously experienced. This is familiar enough: in one sense, people "look tiny" from the top of a tower, but one does not think that they have shrunk; railroad tracks "seem to" converge at the horizon, but one knows they don't. In this space of appearances, profiles, adumbrations, and perspectives, the same objects can be said to grow, shrink, vanish, occlude one another, be "visible-from-here", "inaudible-from-there", and so on. We hold that this perspectival mode of representation is true of both perception and imagination (for both concrete and abstract objects).

Perspectival properties are most obvious in vision (although the extent to which perspectival properties dominate visual awareness is debated [71]), in which the horizon corresponds to the set of vanishing points along lines of perspective, e.g., as parallel railroad tracks converge toward a vanishing point. But as noted, lived space is not reducible to the space of vision. It is a multisensory, integrated phenomenon, incorporating the visual field into a larger space and feeding on sources of spatial sensory sampling other than visual ones. It integrates both perception and imagination into a common space of inference, and is strongly related to embodied cognition, in which prehension and locomotion (i.e., actions) are as much central organising factors as perception or imaginary representations.

One can thus expect that imagination also operates within a perspectival frame of reference, which can be mapped onto the visual field or more generally onto the perceived surrounding space. For instance, when one imagines a cube moving away anywhere around oneself, the imagined cube appears in imagination to become smaller and smaller as it moves farther away, and it eventually vanishes into a point as it approaches "infinity". This geometrical behavior of the cube is associated with a sense of depth.

Since objects vanish near infinity in all possible directions of lived space, lived space, as a perspectival space, appears from inside as bounded by a horizon at infinity, even though the horizon may be set very far in the virtual space of consciousness and remain outside or at the fringes of the scope of attention. We hold that everything present in spatial consciousness is represented in perspective in relation to this movable horizon no matter its relative position in the space around us.

Beyond phenomenological considerations, it is important to emphasize that such a configuration is integral to the mechanism of perspective taking, as it frames all subjective representations of space along specific directions of aiming from specific vantage points. The perspectival structure of the lived space that results is part of the implementation as well as an expression of intentionality in consciousness.

# 2 Model

## 2.1 General Rationale

We engage in cognition and act in ambient space in reference to a 3-dimensional Euclidean space (the ambient physical space may or may not be a 3-dimensional Euclidean space, but assuming a modest realism, it approximates such a space at the scale relevant to most of our practical pursuits). Starting from the vital necessity for biological systems to distinguish between changes in the world resulting from their own motions versus from outside transformations, Poincaré showed the importance of accounting for the relativity of rigid motion between the body and the world in perception and action programming, which can be made on the basis of the group of displacements in Euclidean 3-dimensional space [72]. This particular group appears to structure empirical facts, outside of subjective experience. The action of the group, supported by the Euclidean space, correlates with the behavior of physical space and of fields of matter and light under relative

motion.

However, such physical space does not coincide with lived space as experienced in consciousness. The phenomenological postulates previously reviewed have led us to conclude that the geometry that characterizes lived space is not Euclidean: it is fundamentally perspectival and represents objects in ambient space in a manner that is best explained by projective geometry with a direction of aim and a scope (e.g., "from here" or "over there", or "from here to there"; see [12] for an historical introduction to projective geometry). It follows that the 3-dimensional characteristic phenomenology of the lived space is framed by a 3-dimensional projective space and that the mechanism of perspective taking corresponds to the action of projective transformations. In fact, the group of projective transformations contains the Euclidean group, thus projective geometry embeds Poincaré's transformations.

We are proposing that the generative models used by conscious creatures are necessarily equipped with a projective geometry. Note that the generative models in play cover all aspects of how sensory data are generated. This means that projective geometry constitutes a necessary but not sufficient account of generative models. Having said this, we consider projective geometry to be quite fundamental for conscious processing in the following sense: we associate conscious processing with inference about the consequences of action. This sort of inference is necessarily about the future and therefore acquires a temporal thickness [73]. Crucially, it is the sort of inference that answers questions like "What would happen if I did that?" The fact that we do things in a metric space therefore requires counterfactual inference about the future to conform to a projective geometry. From the perspective of active inference, every (fictive) perspective or view we take on a scene represents a particular movement or possible action, from which we have to select the best. This selection constitutes our intention.

Active inference is a corollary of the free energy principle, which subsumes approximate Bayesian inference and model selection. Common examples of this in the neurosciences include predictive coding [74, 75] and Bayes optimal decision making under Markov decision processes [76].

The PCM incorporates the free energy principle, which mandates that the best perspective is that which results in the smallest free energy or expected surprise. Because surprise is (in information theory) self-information, expected surprise is then entropy or uncertainty. This means that the best among all plausible or allowable perspectives is the perspective that resolves the greatest amount of uncertainty; thereby providing a formal basis for action selection. Perhaps the simplest example of this is deciding where to look next. The selection out of all possible saccadic eye movements can be easily modeled in terms of minimizing expected free energy or uncertainty in terms of salience; either in continuous state space models [77] or discrete state space models [76, 78, 79].

Crucially, expected free energy can always be decomposed into epistemic and pragmatic parts [78]. The uncertainty reducing component is known as Bayesian surprise, epistemic value, information gain or salience [80]. The pragmatic part corresponds to expected surprise under prior beliefs about outcomes; i.e., prior cost. This means that the active inference gracefully subsumes expected utility theory and related (reinforcement) formulations in behavioral psychology.

However, note that neither expected utility theory or reinforcement learning theory are equipped with epistemic or uncertainty resolving aspects  that are a natural consequence of active inference, under the free energy principle. We will not develop the mathematical details behind this aspect of active inference. For details see [10, 78]. Our key point is that the machinery that is necessary to evaluate the expected free energy, under a plausible action or transformation, must necessarily conform to a projective geometry.

The integration of a global projective geometrical level of representation and control into a free energy minimization scheme for multisensory inference in relation to preferences must (we assume) have been shaped through a long process of evolutionary selection  to support the generation of resilient autonomous agents. In this treatment, we have not considered the learning of spatial transformations. In principle, this is usually subsumed under active inference, where the parameters encoding allowable transformations within the generative model are learned so that the

(long-term average) of free energy is minimized. This usually translates into Hebbian or associative plasticity; for both continuous and discrete state space models [81]. There are supraordinate issues which we do not address here that concern the epigenetic specification and learning (i.e. Bayesian model selection) over evolutionary time (i.e., natural selection) [82].

The explanatory work of projective geometry in the context of active inference will be illustrated using a simple example below (with simulations). At present, we just need to note that to select actions and navigate in a metric space, it is necessary to implement a projective geometry to evaluate the consequences of (fictive) narratives. As noted above, an important aspect of this selection rests upon the salience or epistemic affordance associated with each transformation entertained.

## 2.2  Core features of the PCM

This overall framework can be modelled by integrating in a hierarchical manner: (1) a projective geometry engine, which is embedded into (2) an active inference engine; the two being coupled in order to access and evaluate information and control action [1] (figure 1). The projective geometry engine implements a specific kind of geometry, projective geometry, that: (i) is based on perspective taking, (ii) plays a central role in spatial inference, and (iii) is capable of integrating perception, imagination and action-programming into a global operational workspace with seamless relationships between frames of reference. This active inference engine integrates the geometrical model into the larger framework of free energy driven, active inference, which is used as a general model of enactive, embodied cognition [6,8,10]. As part of the overall process of active inference, intentionality, which includes the driving of perspective taking and the motivation of action, is driven by the minimization of free energy, which directly relates to conditional probabilities in memory. These probabilities can encode a variety of expectations, including cognitive and affective preferences, as a function of locations in space. Sensory evidence contributes to the updating of prior beliefs in memory as the system interacts with its surrounding.

Projective geometry implements a specific mode of access to information that is formatted by non-trivial projective transformations. It provides the principles by which a variety of representations or models of the spatial-temporal world, including the body, $R(X,t)$, which are stored in memory, can be accessed and mapped in the workspace from different points of view in an integrative manner. The process is based on the application of projective transformations $T(t)$, which implements perspective taking, such that the subjective experience of space at a given instant corresponds to:

$$S(X,t) = T(t) \circ R(X,t) \ [1]$$

with $\circ$ denoting the application, similar to an operation of composition, of the projective transformation $T$ to the spatial world model $R$, yielding a space equipped with intentional directions in perspective, featuring observation beams and relating subjective perspectives of possible observers.

More specifically, $R(X,t)$ denotes the state at time $t$ of a set $X$ of internal variables that describe the environment, external world (for instance the objects in it, but also the fields on it, such as weather, temperature, light, and so on), and our own body. The state $R$ depends on other variables $Y$, such as sensory inputs associated with sensory data $D$. And $S(X,t)$ denotes another state at time $t$ of $X$ that represents our subjective experience of the external world (objects in it and fields over it), and of our own body. The geometrical hypothesis tells that projective transformations and generalized transformations (for instance the perspective projections, as further described below), acts as a set of transformations in the manifold of states of $X$, sending an unconscious state to a state that is accessible to consciousness.

Moreover, the geometrical hypothesis requires that the full group $G$ of invertible projective transformations, such as changes of projective frames, is acting on the subjective transformations, in the following sense: assume the state $R$ is transformed in $S = T \circ R$ by an element $T$ (that can

be invertible or not), the replacement of $T$ by the conjugated element $T_1 = g_1 \circ T = g_1 T g_1^{-1}$, where $g_1$ is an invertible element of $G$, must satisfy the covariant rule: $g_1 \circ (g_2 \circ T) = (g_1 g) \circ T$.

This means that conscious states are not very different from unconscious ones: they result from taking a point of view or perspective using projective transformations. The choice of the projective operation $T(t)$ transforming $R(X, t)$ into $S(X, t)$ can therefore be assumed to rely on unconscious processing (it can be complex, as we can expect the system to combine several such transformations across representations of the self, other persons and objects).

A more precise symbolization, but also probably more cumbersome at this stage, should take into account two sets of variables, $X$ for variables that are almost conscious, and $Y$ for hidden variables that participate in the definition of $X$ (e.g. related to sensory systems that are far from conscious access). Then the equation should be $S(X, t) = C(T(t), R(X, Y, t))$, satisfying, for any $g$ in $G$, and state $R$ of $X$ and $Y$, the equation $C(g \circ T, g \circ R) = g \circ C(S, R)$. Importantly, the states of variables $X$ and $Y$ have to be understood as estimates of probabilities, which in some cases can be certainties, yielding fixed values.

Such projective mode of access presents strong computational advantages, and can seamlessly relate the spatial frames of perception, imagination and action programming. This model of space constitutes the spatial field of consciousness in our theory.

This spatial field of consciousness (FoC) is controlled by prior beliefs in memory and ongoing sensory evidence to serve active inference through the minimization of free energy. The process results in an algorithm of arbitration and learning that optimizes the projective representation of information (morphological or affective) and the orientation of intentionality (spatial attention and action programming) within the workspace, specifically : (1) adaptive spatial inference (physical and social) to build spatial knowledge in memory (about shapes, structures, behaviors); and (2) 'personal relevance' as mapped onto spatial knowledge in memory. The embodied agent minimizes free energy, as related to the discrepancy between beliefs and evidence, and thus optimizes its representation of spatial information and control of action.

Free energy ($FE$) is a mathematical way of expressing the surprise or improbability of some sensory impressions under a particular (generative) model of how those data were generated. In the PCM, free energy provides an upper bound on surprise or self-information such that:

$$FE > -logP(D|R, P, T, ...) \quad [2]$$

$P$ denotes the parameters of the probability laws of the variables associated with $R$.

This (simple) formulation shows that $FE$ is always greater than surprise. Therefore minimizing free energy (or selecting transformations $T$ that minimize expected free energy) will always reduce surprise (or resolve uncertainty). Here, $D$ are sensory data that are available to the active inference engine. In other words, we associate any possible transformation with a model that, in practice, would become a hypothesis about "what would happen if I did that". As noted above, this enables the expected free energy to be associated with any transformation (or its products); where the products $S$ may also be equipped with prior preferences (i.e., $logP(S)$ of a particular outcome $S$ that has low cost or surprise).

The relationship between free-energy driven active inference and projective geometry in the PCM can be formulated as follows. $FE$ minimization is a general expression of Bayesian inference that incorporates constraints on computational (and statistical) efficiency: as with every Bayesian approach, it relies on a choice of variables to estimate and use internally, and on relations between them. The form of these variables and relations is crucial. A specific geometry furnishes a way to make precise and organize this form; in fact a specific geometry defines particular observations to be made and particular rules prescribing how the observations and priors have to be analyzed. Here we can refer to Poincaré's analysis of the most convenient geometry for managing the ambiguous relations of individuals with the external world. In the specific case of consciousness, the most convenient geometry is a projective one, notably because projective geometry is the natural geometry of arrays of directions from any point to all other points; it relates "me" observing and "me" being observed together, taking care of several frames of observation from places to places.

Moreover, free energy minimization offers a highly flexible, fully computable framework that can be associated with concrete algorithms for actually performing quantitative predictions and advanced simulations of agents implementing the PCM, including in guiding perception, action and decisions. $FE$ offers a solution to the problem of implementation in addition to having a strong theoretical back story. The behavior of the model reproduces psychological descriptions invoking states of affairs of personal relevance (rewards) and utility. It allows us to understand and manipulate psychological dimensions computationally, based on a general principle of optimization of resilience, in order to test hypotheses about the role of cognitive and affective processes in behavior. It also allows us to make connections with brain function and more generally shows how geometry can be compatible with probabilistic algorithms for learning and decision

## 2.3   Incorporation of models of appraisal and motivation

The free energy framework combined with projective geometry conveniently enables us to explicitly incorporate appraisal and motivation theory by setting prior beliefs in memory and rules of approach and avoidance [28, 31, 37]. We can thus specify the mechanics of embodied consciousness and condition perception, imagination and action, based on affective and motivational models, which can be related to given adaptive goals of survival and well-being, and compared based on simulation outcomes. We place emphasis on these affective aspects because they are essential psychological and ethological concepts that a unified computational framework proposed as a model of embodied consciousness should be able to account for.

We can interpret the free energy framework, under some specifications, as implementing an intuitive notion of 'personal relevance'. The notion subsumes innate and acquired abstract and concrete 'values' or preferences, yielding cognitive and affective saliencies, as built into the prior beliefs about the world of a given system. In this context, personal relevance and motivational valence are scored by the expected free energy under a particular transformation (entertained by the projective geometry within the generative model). As noted above, the motivational valence or value is encoded in terms of prior beliefs about (counterfactual) outcomes $S$, while the personal relevance of those outcomes corresponds to the salience or epistemic value of reducing uncertainty about how these outcomes are obtained; i.e., epistemic affordance. Clearly, the salience or epistemic part of expected free energy is inherently personal, because it corresponds to the particular viewpoint or perspective implicit in each possible transformation. We operationalize personal relevance as the negative logarithm of a field of conditional probabilities that represent prior beliefs; namely, events with a given affective and motivational valence for the agent or 'value' will occur across a given portion of space. This constitutes a belief system or hypotheses about the world stored in the agent's generative model. Through active inference the agent confronts its beliefs with a given 'factual' world, which corresponds to sensory evidence that can be 'experienced' by the agent as a result of its sensory coupling with the world. Beyond appraisal and motivation, notions of motor commands and sensorimotor forward and inverse modeling can be naturally incorporated into the free energy framework, and become integral to the design and understanding of specific systems performing active inference. In active inference, motor commands become descending predictions of proprioceptive (or interoceptive) signals, which elicit motor (or autonomic) reflexes [83–85]. By the same token, corollary discharge becomes the corresponding descending predictions from the generative model to exteroceptive modalities (e.g., what things would look like if I moved there) [79]. The PCM can thus perform active inference within the real (e.g. through robotics) or simulated (e.g. through virtual reality) environments that define a space of interaction and a world model.

## 2.4   The Field of Consciousness in the PCM

Thus the minimization of free energy drives perspective taking by setting the parameters of the projective transformation $T$ based on a combination of cognitive and affective parameters, related to the expectations of the system. This process plays a central role in intentionality and the

motivation of action. For instance, typically, we want to move away from that scary thing over there we believe to be dangerous, we want to move closer to one who attracts us, we stay put to continue what we are enjoying as long as other motives do not interfere. In the process we sample sensory evidence and revise our prior beliefs accordingly, and consciousness helps to transform implicit computations into explicit ones that are sensitive to observations.

In the PCM framework, under the minimization of free energy, the $FoC$ is not only a spatial field, but also an integrated "force field". The agent is governed by a dynamical system that drives the mechanics of its embodied consciousness through perspective taking across virtual and real spaces, to maximize spatial knowledge and personal relevance and control action. We note that our model does not offer an explanation as to why the force field is phenomenally conscious; it limits itself to addressing why consciousness has the shape and dynamics of such force fields. The spatial gradient of free energy corresponds to the spatial gradient of preferences.

The overall $FoC$ at each instant, including the agent's intentionality (i.e., the direction of its aim), is driven by the dynamic interplay of ongoing multisensory information, memory and affective biases, and serves to appraise and decide on the adequate course of action, based on accumulated, anticipated and observed information. By combining spatial and affective priors under the free energy principle, we can specify the principle of action that governs embodied intentionality and actual action in our model, in a way for instance that minimizes the likelihood of unexpected and adverse events and maximizes positive outcomes for the embodied system. The framework allows us to encode space (including shapes) and affective valuation together, so that active inference relies on an "aesthetic" process so-to-speak, which is consistent with the strong link between spatial and emotional memory in neuroscience [7].

When we associate transformations with an attentional beam, we are appealing to the notion of attentional salience and to the epistemic value entailed by minimizing expected free energy [79]. This should not be confused with the perceptual attention to ascending sensory streams; usually cast in terms of precision weighting in predictive coding formulations of active inference [86]. This means the attentional beam corresponds to the epistemic affordance that characterizes a particular transformation and ensuing perspective on the world.

## 2.5   Illustration of the PCM

figure 1 illustrates the general properties of the PCM from a phenomenological standpoint. The subjective experience of space and its contents as perceived (continuous funnels on the figure) or imagined (dashed funnels), i.e., lived space $S$ for short, are framed in, and experienced as, a 3-dimensional projective space (blue spheres) in perspective, through a mechanism of perspective taking. Personal relevance (as associated with sensory qualities and feelings in conscious experience) is represented in the colors on the manikins. Perspective taking is governed by projective transformation $T$, which is applied to the current world-model $R$ in memory, under the influence of executive functions, emotions, and ongoing multisensory integration. The minimization of free energy drives the dynamics of the system, including the sequence of perceived or imagined projective scenes (i.e., little mind movies) that plays out in its consciousness in an emotionally charged context. The overall process is implemented through the coupling of (a) a real-time projective geometry engine, performing subjective framing and rendering, and (b) an active inference engine, appraising perceived and imagined projected situations based on internal models and prior beliefs. The body is represented and experienced as a central reference in the space, and the visual field is only a part of lived space.

**figure 1.   Overall sketch of the model.** A. Lived space $S$ (blue spheres): subjective experience of space and its contents as perceived (continuous funnels on the figure) or imagined (dashed funnels). It is framed in, and experienced as, a 3-dimensional projective representation in perspective of a world model $R(X,t)$ associated with a distribution of free energy (FE) in space, related to cognitive and affective personal relevance (colors on the manikins and door, and disk

with color gradient). Perspective taking and the selection of corresponding projective transformations $T$ are governed by the gradient of free energy across space $\delta$FE. The diagram with the arrows in circle represents the possible transitions of state: form perception to imagination and action, for the Field of Consciousness (FoC), which is partially driven by sensory evidence when it is focused on perception of the local environment, and entirely driven by prior beliefs in memory and simulation capabilities when it is focused on imagination of non-local spaces (e.g. when we imagine ourselves at home while we are at work, in the present, past or future), which cannot be based on direct sensory evidence. —colorredThe $FE$ defining the optimal perspective at a given instant is globally minimized, through cycles of perception, imagination and action, across time. The bar graph represents FE, as a function of time and of the different perspectives and processing modes (P: perception versus I: (non-local) imagination) adopted by the system (see text). B. Overall diagram of the active inference engine, embedding the projective geometry engine for subjective rendering, confronting prior beliefs with multisensory information and driving learning and action. The overall process minimizes free energy.

An hypothetical scenario can be used to illustrate the basic operation of our model (see figure 1). A loud explosive noise from behind captures the attention of a man governed by the model (blue manikin). He stands in a living room with an open door in the back (as an accessible knowledge $R(X)$ stored in his memory)(figure 1-a left-tier)($t0 - t1$). Alarmed by this stressor (free energy is increased), he starts imagining (the projective geometry engine is solicited) that he might be under attack (he has a strongly biased prior belief that one can now be attacked anytime anywhere from behind, based on what he recently saw on TV and heard on the radio repeatedly). This anticipated attack manifests in his momentary consciousness as a mental projection $S(X, t2) = T(t2) \circ R(X, t2)$ of a spatial representation $R(X, t2)$ of the imagined situation, which is construed from a certain imaginary standpoint (embedded in $T(t2)$). The man does not confuse this counter-factual imaginary projection with a perception (he is not hallucinating): we assume that as long as perception is tied to direct sensory evidence ones believes that one knows what one seems to be perceiving, and thus also knows what one does not seem to be perceiving, in particular when information is based on long-term memory access, so that the counterfactual scenario imagined is known not to be among the situations currently perceived. As part of a process of active inference, the projection nevertheless demands action in order to confront predictions with sensory evidence. The man expects, when he will have turned around, to see another man (red manikin) pointing a loaded gun at his head, in front of the only door ($t3$). He slowly starts to turn around (body action and $Rot(theta)$), hindered by a sharply growing fear (free energy spikes very high)(figure 1-a right-tier). He imagines a possible more positive outcome, preparing accordingly for action: to run away through the door and escape the danger (free energy is reduced slightly by anticipating this scenario)($t4$). But instead of this dire outcome, as he now stands in front of the door (following body action)($t5$), he sees the door but no gunman in front. There is a strong smell of plastic smoke, the TV is on the ground broken. The cat (green figure) jumps out of nowhere and runs away, which he sees first, then only hears, and all of which he perceives as one unique scene in his global consciousness of space, $S(X, t5) = T(t5) \circ R(X, t5)$. He now understands that the cat must have managed to make the TV fall to the floor, causing the loud explosive sound. Relieved ($FE$ is lowered), he now smiles at his mistaken panic-attack, staring emptily at the door, updating his prior beliefs in memory (i.e., reappraising) towards the greater likelihood of one's cat messing around than of being attacked at home.

## 2.6  3-dimensional projective structure of the FoC

Here we further specify the geometry embedded in the PCM as a model of the subjective experience of embodied consciousness that defines the FoC.

### 2.6.1 The FoC as a 3-dimensional projective space

Technically, a projective $n$-space $\mathbb{P}^n$ can be defined on fields of numbers $\mathbb{K}$, as the collection of all lines that go through the origin of the vector space $\widetilde{V} = \mathbb{K}^{(n+1)}$. Thus by definition, $\mathbb{P}^n(\mathbb{K}) := (\mathbb{K}^{(n+1)} \setminus 0)/\sim$, with $\setminus 0$ expressing through a set subtraction symbol $\setminus$ that the origin of the space is excluded from the projection. This is because if 0 were included, all the points would become equivalent. The symbol $/\sim$, expresses the equivalence relation which relates points along the same vector lines. The vector space $\widetilde{V}$ has one more dimension than the projective space $\mathbb{P}(\widetilde{V})$. In what follows we only consider the case of the field $\mathbb{R}$ of real numbers, which is a model of continuity. Moreover we will mainly limit ourselves to $n = 3$.

In a projective space, the projective subspaces are defined as the sets of lines that belong to vector subspaces of $\mathbb{R}^{(n+1)}$. For instance, a projective line is the set of lines through 0 in a 2D plane, a projective plane is the set of lines through 0 in a 3D space, and so on. Hyperplanes correspond to subspaces of codimension one. When a hyperplane is selected, chosen with an equation that defines it in $\mathbb{R}^{(n+1)}$, its complementary subset in the projective space $\mathbb{P}^n$ has the structure of an affine space $\mathbb{E}^n$ (associated with the corresponding vector subspace of dimension $n$ as a space of translations). Then parallel lines in $E^n$ converge toward vanishing points $p^\infty$ at infinity, the set of which defines the horizon hyperplane.

### 2.6.2 The FoC under the action of projective transformations

Projective spaces and representations in projective spaces are ruled by the action of projective transformations from the Projective Linear Group $PGL(\widetilde{V})$. These transformations govern global changes of perspective taking. Key parameters of the generative models that we consider thus become projective transformations $T$. One practical advantage is that linear algebra and matrix operations can be used to represent the whole process.

For a projective space in 3 dimensions, the group corresponds to the set of all $4 \times 4$ invertible transformation matrices $M$, which transform points in the ambient space (represented by homogeneous coordinates), to points in projective space, through changes in perspective taking in a 4-dimensional space ($M$ is a subset of the possible transformation matrices $T$, which include non-invertible transformations). The group includes all affine transformations (rotations, translations, shearing and scaling) and projective transformations that define direction of aiming and perspective in the space while preserving relations of incidence, without necessarily preserving the plane at infinity. The reciprocal switch between points of views is governed by simple projective parameters. All possible projections are thus related via a set of transformation matrices, which are invertible and can be estimated. In this context, $T$ denotes such an invertible transformation $g$, which corresponds to an invertible matrix $M_g$ in $PGL(\widetilde{V})$.

It is also important for the model to include *generalized projective transformations* in the projective space $\mathbb{P}$. They correspond to non-invertible $4 \times 4$ matrices different from 0, or non-invertible linear applications of $\mathbb{R}^4$. A particular case is the projection from a point $O$ in $\mathbb{P}$; in this case the kernel of the associated linear map is the vectorial line corresponding to $O$. From a plane to another plane it is an ordinary *perspective*, the most usual for drawing, that projects the horizontal plane, including its horizon at infinity, to the vertical plane in front of the implicit point of view.

A fundamental result of projective geometry is that all projective transformations from a given plane to another plane, i.e., a transformation that preserves pencils of lines, sending lines to lines, are obtained by composition of usual perspective transformations centered at variable points [87].

## 2.7 Relations between FE, projective geometry and the optimisation of preferences in the PCM

The projective geometry engine serves as a unified level of representation and control for the integrative process of active inference: (1) about the causes of limited sensations and the uncertain

consequences of action [10, 88, 89]; and (2) about the gradient of personal relevance in space in relation to embodied cognition and multiple dimensions of appraisal [7, 8, 26, 28, 31, 37, 90].

It provides the agent with a method of situated representation of space based on which the process of active inference can progressively build an integrative non-local representation $R$ of the Euclidean world, which can be associated with indices of personal relevance, or likelihoods of affectively valenced events in memory based on the $FE$ framework. $R$ and its valuation are not directly accessed in consciousness as such, as they correspond to the entire relational knowledge and beliefs potentially available to the system (or a subset of it). Conscious access is always a contextual, situated process, which builds upon perceptual inference [22], using perspective taking to support generalizations and action programming.

There is a one-to-one mapping in the PCM between the elements of $R$ in space and the valuations in memory of cognitive and affective personal relevance across levels of appraisal. The projective representation $S$ acts as a spatial filter, which frames both the world model $R$ and the associated mapping of preferences encoded in the prior beliefs in memory. This provides a principle of driving for perspective taking across perception, imagination and action that orients consciousness according to the expected spatial distribution of preferences. Through $FE$, the process is tuned to be sensitive to surprise, that is, the discrepancies between expectations in memory, e.g., as accessed through conscious imagination, and the actual state of affairs, as perceived. The balance between discrepancy and confirmation contributes to drive the parameter of "free energy".

As a generative model is learned by revision of priors or "Bayesian updating" under the driving of free energy, minimizing free energy corresponds to optimally approximating the true posterior probability of the state of the effectively observed world, both in terms of content and personal relevance in our model. Free energy minimization thus operates as a biased operator of knowledge generation (biased by preferences) when associated with a mechanism for updating prior beliefs.

In this general framework, "active" inference (which implies actual or imaginary motion) is a method of optimization for the collection of observations (e.g., from sensory data $D$), aimed at updating previously estimated rules for hidden variables (e.g., the world model $R$) and parameters (e.g., valuations of locations and objects). It selects observations which are best predicted [10, 91]. In some sense, this is a reverse optimization algorithm, or an optimal conditioning for further likelihood maximization. Importantly, in this framework, violations of expectation can be resolved through two complementary strategies: (i) to a revision of prior beliefs (e.g., internal models of the world) or reappraisal (depending on learning rules), in order to ameliorate prediction by reshaping memory; or (ii) to adaptive actions aimed at confirming prior expectations, thus enacting biases of the system (e.g., which can lead sometimes, in relation to learning parameters, to ignoring part of the world because it does not match prior perspectives).

As a result, the system maintains a dynamical balance between, on the one hand, the use of fixed prior beliefs to resolve uncertain situations, and on the other hand, the update of prior beliefs based on new evidence that violates expectations or preferences. The arbitration of this balance governs the choice between strategies of cognitive reappraisal or of active approach or avoidance. It can be expected to vary broadly depending on the design, experience, developmental emotional history, and current situation of the system. Nevertheless, in nature, the situatedness, autonomy, embodiment, and limited processing power, memory and energy of embodied autonomous systems, as well as the relative stability of their environments, can be expected to make biases or prior beliefs carry a heavy weight in inference and choice, and to impose a slow and reluctant process of updating, in many contexts. In other contexts (like the resolution of everyday perceptual uncertainty) the updating is normally rapid and eager.

Free energy thus functions as a control parameter, which arbiters and motivates, like a tension to reduce the (costly) implementation of changes in the embodied agent's states, through cognitive reappraisal or adaptive action, as they appear necessary to reduce past, current or anticipated violations of expectations. In our model, free energy fuels and puts stress on the workload of the inference engine and drives learning. Minimizing free energy amounts to maximizing the predictive

power of internal models in memory (including prior goals and preferences) in relation to sensory information. Because of its intensive dimension and driving role in active inference, free energy can be interpreted from a psychological standpoint, as contributing to attentional drives, emotions and the motivation of reappraisal or action.

The concept of free energy straightforwardly subsumes key features of emotion, as a motivational force for reappraisal and action. It can be expected to relate to a feeling of anxiety when it increases as the embodied agent misses its objectives, motivating the agent to act, or paralyzing it but forcing it to reappraise its predicament; or to feelings of relief as $FE$ decreases when everything turns out to be fine, perhaps as initially expected or hoped in imagination [9, 92, 93] (see figure 1 and accompanying scenario).

One key specification on the free energy framework in the PCM is that it drives intentionality (the "aiming at" function in the conscious field) and the corresponding perspective taking (the specific projective transformation $T$): gradients of free energy across space, as mapped in memory and evidenced through sensory sampling, directly drive the selection of $T$ and of the corresponding possible direction of perception and action. Free energy minimization thus performs an optimization of intentionality in the conscious field, which behaves similarly to the needle of a compass aligning with the magnetic field. The algorithm selects the perspective (and corresponding projective transformation) that minimizes the total amount of free energy associated with objects under the scope of the current field of consciousness, acting like a magnifying device, tuning down everything that is not directly under its lens (see [88] for a related discussion at a different level of integration). This results in the optimization of personal relevance according to specific preferences: e.g., we look towards the fridge when we are hungry, and listen to the people behind that are talking about us. For instance, a high likelihood of positive personal relevance (e.g., pleasantness and safety) in consciousness will be associated with a low level of free energy, whereas a low likelihood of positive personal relevance will be associated with a high level of free energy (or stress). These dynamics operate across perception, imagination and action. Free energy itself is not directly accessible to consciousness, and its minimization process relies on unconscious activities. However, part of the knowledge that evolves through the optimization process becomes conscious, and this part necessarily involves the embedding in a lived workspace.

## 2.8   Projective geometry as an optimal tool for embodied active inference

From the standpoint of natural selection, given the central role of spatial inference in our model, the phenomenology of consciousness is expected to present strong and specific computational advantages, especially for spatial inference itself. Projective geometry implements a framework that enables intentionality, perspective taking and action planning in the system, in a manner that can be driven according to the globally optimal principles of free energy minimization. But projective geometry in and of itself is also optimal for a variety of problems pertaining to situated spatial inference. In other words it is the best geometrical model to perform spatial inference from situated perspectives [13].

In the context of active inference, the inference engine must progressively update the representations $R$ stored in memory, by building a more and more comprehensive representation of the spatial and temporal world, based on a limited amount of situated observations, and a set of preferences. In the process, the engine selects situated perspectives, (i.e., transformations $T$), which optimize the acquisition of new multisensory spatial information $D$, based on available prior knowledge, and under the driving of some contextual mapping of personal relevance. For example, I'm curious, I see a door in front of me, I don't know yet what is behind it; I can perhaps expect a small room that I represent in my mind from different perspectives compatible with my expectations, which will be confronted with the actual structure of the room soon by my opening the door and entering in order to update my world model $R$ accordingly. Such a problem of spatial inference can be optimally solved using projective geometry as a computational method.

Projective geometry is optimal for forward and inverse modeling in space, because it enables the manipulation of a rich set of frames of reference.

Thus by combining a given set of perspectives from perception using projective geometry, a system can progressively infer a broader and more precise representation of the 3-dimensional world, in a manner that is independent of the particular point of view and that can be stored in memory for future retrieval. Such spatial inferences can be carried out based on epipolar geometry, which is part of projective geometry, and can be formulated as a multiview process relying on the theory of abstract cameras [13]. It is computationally tractable, as it can be approximated with linear algebra. It is reliable, as the shape and distribution of objects in 3-dimensional space can always be inferred through projective transformations and a small set of additional priors. The approach is analogous to the methods used for the 3-dimensional mapping of objects, e.g., from sets of 2-dimensional pictures [13], as used in architecture, virtual reality, engineering and in industry (more generally, projective geometry is the basis of Computer Graphic Imaging (CGI) and computer vision).

Moreover, with projective geometry, the frames can link complex perceptions with complex predictions and actions in a seamless way in order to explore and build a representation of the environment. (Higher dimensional projective spaces might need to be considered to account for the multiple dimensions of consciousness [consistent with multimodal sensorimotor integration], but the essential characteristics of the phenomenology we focus on here, the 3-dimensional lived space, can be understood by considering 3-dimensional projective space).

Given estimated sensory data $D(t1)$ and internal model $R$ representing the state of objects in space, it is possible to estimate multiple possible perspectives on the current world-model within projective spaces and to estimate the projective transformations induced by displacements of the frame from one place to another one in the space. This estimation produces the expected values of the sensory data $D(t2)$ after displacement. The system can then use a set of real displacements and compare the new sensory changes with the predicted observations expected from new points of views. The whole process relies on the general role of geometrical invariance in the preparation of action and the prediction of the effect of actions on perceptions [94]. These mathematical properties allow the system to run a continuous optimization process that updates the hidden rules and the internal model based on spatial cognition.

## 2.9 From the PCM to an analytic definition of perception, imagination, and action

The question of the distinction between perception and imagination has been a classic topic of debate in philosophy, psychology, and phenomenology [95–98]. A model of consciousness should be able to account for their distinctions and relations, for the unity of intentionality and lived space across these forms of consciousness, and for how they relate to action. The PCM suggests a way, in connection to active inference, of understanding this in a formal manner, which we use later on to account for psychological phenomena.

### 2.9.1 Perception

Perception can be associated with a specific subset of generalized projective transformations, the perception spectrum for short (these transformations do not belong to the proper projective linear group and correspond to linear projectors parallel to a line on a hyperplane equipped with a Cartesian frame, see the section on the workspace under the action of projective transformations). (1) The point of view is located close to or within the space occupied by the head of the body in an egocentric mode. In normal conditions, this means that the projection operates from an embodied first-person perspective, with the real body as a physical referential, and the virtual body being correspondingly centered. Such setup provides optimal conditions for construing perceptual evidence, as it registers the center of projection in space with respect to the locations and directions of the sampling of space that is performed by actual sensory interfaces (e.g., the eyes,

ears and the body envelop with the skin) based on sensory data $D$, which incorporate intrinsic constraints of orientation and calibration related to their source organs. (2) The frame is oriented in relation to the body referential and main axes: the aiming or "look-at" projective vector tends to point frontward by default, but this may be modulated by attention, and the up-vector is normally defined from inertial information (e.g., the direction of gravity sensed by the vestibular system). (3) The projection parameters (e.g., field of view, perspective) are set to a specific calibration range (e.g., encompassing the visual field).

At each cycle of perceptual inference, the system must identify the projective transformation $T$ that will best fit sensory data $D$, given its sensorimotor state. Forward models of sensory organs and processing are expected to serve as priors, in order to reduce the search space for $T$. With such priors, perceptual access can assume a fixed subset of possible projective transformations $T$ to explain sensory data $D$, as sampled by sensory organs and conditioned by internal priors $R$, and can quickly focus attention on personally relevant information. In other words, perceptual access, which must be as close as possible to real-time processing, operates based on a regularization principle, in order to solve the inverse problem of representing the 3-dimensional spatial causes of $D$ in the workspace, and framing them according to personal relevance. Only transformations $T^P$ from the perception spectrum will normally be selected for forward modeling. (We note that our notion of "sensory data" here is neutral with respect to various competing theories of perception. We are not committing ourselves to classical sense-datum theories. See [99] for some discussion. All we need here is the idea that in one way or another the sensory organs provide an independent source of input and correction for the continually updated world model we describe.)

### 2.9.2 Imagination

The imagination can be associated with the full group of projective transformations $PGL(P)$, and generalized projective transformation as well, the imagination group for short. It includes all possible compositions of perspectives from higher dimensional spaces, inducing changes of frames in the space, including the "God's eye point of view" near infinity, which offers a view of almost the entire space, from any possible direction. It can be applied to conscious access in the context of a representation of a world $R$ that can be realistic or a fantasy world.

In the PCM, the imagination serves as a space for inference, based on anticipations and simulations of the self vis--vis itself and the world. It is used: (i) to assess past, present, and future, physical and social scenarios according to personal relevance; (ii) to learn from experience by updating internal models in memory accordingly; and (iii) to program future actions based on the simulation of multiple perspective changes.

The imagination group can be seen as the perception spectrum of external consciousnesses or points of view on the space. It notably includes transformations of the observer itself, in particular of its location in and outlook on the space. It can always be used to infer how a scenario would look under perceptual constraints, from an egocentric, first-person perspective. Transformation matrices can be computed and applied in order to compare perceptual situations (e.g., after motion), with imaginary situations (e.g., as anticipation of perception before motion).

### 2.9.3 Perception versus imagination: computational complementarity

**Computational similarity: implications for inference** .

In this framework, imagination and perception appear fundamentally related through projective geometry. Mathematically, they correspond respectively to the full projective group and to the set of projections from points to planes. They share projective transformations to seamlessly interact across the projective workspace, via the operation of perspective taking. This relationship is essential in the process of active inference through which the agent accumulates spatial knowledge about the self and the world. Actual perceptions can be compared to anticipations of perceptions based on imagination, in order to revise prior beliefs and further minimize free energy.

**Computational simplicity: an attentional filter** .

The projective setup makes comparing perception and imagination computationally simple. Sensorimotor contingencies can be used for the calibration of a consistent metric in the projective space. This can be done based on a unit spherical geometry in the 4-dimensional real vector space (such a geometry is well adapted for vision, in which distances are measured with angles). Alternatively, it can be done by choosing a plane at infinity and a Euclidean distance in affine space (which is adapted to voluntary motion of the body).

Projective spaces then offer natural error metrics, which are based on an intuitive space of objects in perspective that can directly be compared, for instance using measures of distances between shapes in perspective (e.g., the point-to-point difference between the expected smile and the blank face one receives). Because the space is in perspective, carrying out the comparison in projective spaces automatically down-weights the importance of information that is located farther away in relation to the attentional scope: the precision or sensitivity to spatial discrepancies or noise is larger for near objects than far objects. The scope of the projection thus acts like a lens and an attentional filter, rendering it a powerful computational tool.

**Differences: role of imagination in expanding operational knowledge and energy saving** .

Imagination and perception are nevertheless clearly distinguished in the model. The respective weights of sensory data $D$ versus data from episodic and semantic memory (as opposed to sensory buffers), in the current definition of the accessed $R$, as well as that of different prior beliefs about possible transformations $T$ for perspective taking, form multiple levels of constraint for the distinction between perception and imagination. In perception, $D$ drives the inference. In imagination, contrary to perception, the agent builds its representations of a local world more based on data from memory and internal simulations than on online sensory data $D$. Consistent with these distinctions, in sleep, the precision (i.e., weight) afforded to sensory data is attenuated through neuromodulatory mechanisms, permitting a fictive exploration of projective geometry that can take on a fantastical aspect as we dream, all the more as intentionality, perspective taking and actions are driven by a purely imaginary field of personal relevance [100].

Another important difference between perception and imagination is that imagination, contrary to perception, can be deployed in a manner that is largely free from physical and inertial constraints (e.g., one can jump to the Moon in a fraction of a second), including constraints on the calibration of the frame of reference itself. It is also true about social constraints that can be released in imagination. In contrast, perception must submit to the independent order of the physics of inertia in the world within which it is operating. The imagination allows the agent to explore multiple possible worlds and actions without having to visit them and act (i.e., with less consequences). It is a framework for the optimization of action plans, which will de facto minimize energy expenditure.

### 2.9.4   Action

Embodied autonomous agents need to act based upon their inferences. Programming or predicting action requires the extraction of parameters of motion. According to the model, conscious access carries out a projective mapping $S = T \circ R$ of a representation $R$ of the local world (including the body), according to cognitive and affective preferences. This representation is itself a mapping of a locally Euclidean, ambient affine space $\mathbb{E}^3$, corresponding to the physical space. The projective space $\mathbb{P}$ and its affine subspace $\mathbb{E}^3$ are linked by geometry. Passing from one to the other only requires linear algebra and adequate calibrations of the respective frames of reference (e.g., between the distance I cover in the real world and the distance I seem to cover in projective space). Such calibration needs to be actively performed by dedicated systems (e.g., the cerebellum)(figure 3).

The projective linear group $\mathbb{P}$ contains all the Euclidean transformations of the affine subspace $\mathbb{E}^3$. With proper calibration, transformation matrices embed the main parameters of displacements in $\mathbb{E}^3$: translations, rotations, and twists. The parameters of real motion can thus be estimated for

action from imagination-based projections of the motion, with a corresponding series of projective transformations anticipating perception. The system can always predict based on its internal models, how the environment will appear from a given position, by choosing the corresponding transformation $T$ at any point in time, and by using it to extract the parameters of motion. The system will thus be capable of imagining optimal navigation paths based on its current representation of the local world $R$.

The classical concept of efference copy [101] denotes an internal sensorimotor prediction accompanying an order of movement, aimed at preparing the organism for a new expected state, and serving as a reference for control. Efference copy informs the central nervous system about the active contribution of the subject in the perceptual process. Part of the concept can be reformulated projectively. The projective workspace acts like a space of prediction. It uses a set of projective transformations $T^I$ to predict how $R$ should appear after motion under a new standpoint. This prediction corresponds to a projective notion of efference copy in our model. After effective motion, such an efference copy embodied in imagined motion can be directly compared to perception-based projections in the projective workspace. Furthermore, transformation matrices, accounting for the difference between actual and anticipated perception, can be computed based on projective geometry, and used in order to compensate for the errors of prediction in action consequences, e.g., in relation to un-modeled perturbations in motion. Note that in the programming of action the minimization concerns the expected free energy, not the actual free energy. Alternatively, the predicted sensory data under a posterior belief about a transformation could be realized automatically through motor reflexes (or saccadic eye movements; see above). This is the tenet of active inference in which descending predictions or efference copies provide references or set points for (reflexive) action.

## 2.10 The imagination: an algorithm for escaping local minima of free energy in the FoC

.

The arbitration of the balance between remote imagination (in time or space) versus local perception is critical for adaptation. In the PCM the FoC shifts between the two processes. This capacity to switch between local and non-local processing functions as an optimization algorithm: when the local FoC does not yield a better minimization of free energy, i.e., increases preferences, non-local imagination can take over.

The stability of free energy across time can be used as a higher-order control parameter to motivate a switch from a local perception mode to a remote imagination mode during which the agent scans its memory (its current prior belief map) to identify occluded or distant locations at which personal relevance would be better optimized. The switch can be implemented based on the monitoring of the stability of free energy across time. Stable free energy throughout exploration of the local environment in the FoC suggests that no further optimization of the FoC can be expected locally. When the local FoC and its immediately reachable surroundings do not provide the agent with access to locations that increase preferences, the best strategy for the embodied agent is to search beyond its situated perspective by projecting itself away. When in a certain range of stability of $FE$, the system projects itself at remote locations from different perspectives, searching for the direction of maximal utility, i.e., minimal free energy, and programs action accordingly, based on its belief in memory. It is an empirical fact that imagination is often triggered by a lack of change in the environment, e.g., during low-level repetitive tasks, which induce mind-wandering [66].

In the PCM, when free energy becomes too stable (e.g., as measured with first-derivative stability analysis), the system may be trapped in a local minimum with no better perspectives. For instance, when the embodied agent cannot find a location to optimize its situation, i.e., FE does not decrease, the system might wander around its immediate location aimlessly, which, interpreted psychologically, may translate into feelings of panic, hopelessness or boredom, depending on the

level of baseline free energy. More generally, we expect the dynamics of free energy to relate to basic and social emotions such as surprise, fear, anxiety, relief, joy, sadness, shame and pride, in the driving of perception, imagination and action.

In line with these ideas, free energy derivatives (i.e., the changes in free energy as a function of time) have been proposed to be related to affective valence, and to control through their dynamics, the elicitation, expression and transition of various affective and motivational states [92, 93].

In the PCM, the imagination thus belongs to a class of optimization algorithms that incorporates mechanisms to mitigate the risk of solutions that are trapped in a local minimum of energy. Through active inference, and the ensuing balance between exploration and exploitation, the system further maps the free energy landscape, using actions and updating prior beliefs, in order to further optimize its FoC. Free energy (and its derivatives) can be thought of as a state or phase space shaped by a landscape of conditional probabilities that describes where perception, imagination and action are most likely to rest or move to and from. The ongoing intentional state of aiming ensuing from FE minimization projects the agent across this space, making it move towards states, or attractors, further minimizing free energy, either through cycles of action and perception or through imagination. This renders the capacity for resilience of the system globally optimal.

## 2.11 The body model as the central reference of the space and its projective avatars

The PCM as presented so far generates a workspace embedding a mechanism of conscious access to spatial information in perspective, based on perspective taking, and driven by the minimization of free energy in relation to encoded personal relevance in space. In embodied systems, consciousness as a level of representation is also characterized by the pervasive presence of a 3-dimensional representation of the body (see postulates above), which is itself a space of high personal relevance.

### 2.11.1 The body as a projection in conscious access

Applying our model, we hold that the body at stake in subjective experience is of a projective nature: it is accessed in consciousness projectively, like other objects in space, in perception, imagination and action. This hypothesis implies that there is a body model (or schema) $B$ that is embedded in the current representation $R$ of the surrounding world, serving as a mobile referential in the space [11, 28, 102]. It also implies that the body model and its dynamics are projectively mapped into the conscious workspace based on the application of projective transformations $T$, to control perspective taking, according to $S(B) = T \circ B$. (Technical remark: here, for clarity, we separate $B$ from $R$, but we could introduce the part $X_B$ of internal variable $X$ that concerns the body, and keep $R$ for denoting the complete state (probabilistic law) of all variables, including $X_B$.)

In normal perception, the set of projective transformations used for bodily conscious access will correspond to the perception spectrum, yielding an egocentric, first-person perspective access to the body in space. In imagination and action, the full projective group can be used for perspective taking on a 3-dimensional representation of the body. In abnormal conditions, many dissociations and personally relevant violations of expectations can be envisioned between the real body and the projected-perceived and projected-imagined bodies.

The body model nevertheless plays the role of an essential reference in the space, irrespective of perspective taking: we always know where to find it in normal conditions, no matter how unfamiliar our surrounding environment may become. It is animated and updated in real-time in consciousness based on multisensory information $D$.

### 2.11.2 The projective body, emotion and feeling

The body model and its internal structures define at each instant portions of space, which are imbued with personal relevance: it is the core space for the experience of pleasure and pain, and

for the interoceptive, homeostatic monitoring of the health and well-being of the agent [28].

As an integrated representation and a critical asset for the embodied system, the projective body serves utility and adaptation, in both physical and social contexts. It is the object of constant affective appraisals [30, 90, 90, 103], and is highly valuated. As part of core consciousness, it is a space of experience, which, through interoceptive sensations, emotion and feeling notably, often reflect the consequences and discrepancies of actual and imagined actions [28]

But we can further wonder how feelings themselves as subjective experiences could be conceived in this projective framework. We suggest that bodily feelings can be in part characterized by a spatial phenomenology. We experience our body as a sounding board [104], and bodily feelings are perceived as partly localized in the body compartment, in relation to specific organs and systems (e.g., the heart, the integumentary system). Feelings are also experienced as representations of motor tendencies, with anticipated postural and behavioral changes, which also yield spatial representations.

We thus can conceive of bodily feelings as projective representations of the body in space, in part or as a whole, undergoing transformations, including representations of the internal state of the body, ultimately driven by the dynamics of free energy. More specifically, these representations can be driven by (i) ongoing real-time sensorimotor information, (ii) memories, and (iii) anticipatory imagination. The nature of the relationships between such mental processes and the gamut of actual physiological processes in the body hic et nunc remains an open question.

At a higher level of integration, the projective workspace can also be thought of as an internal environment for building, monitoring and housekeeping a self-image in relation to social cognition, in which the body image as appearing under certain perspectives plays a central role, and is constantly reappraised in terms of personal relevance [26, 105–109].

# 3  Results

In this section, we apply the PC Model to a variety of behaviors and psychological phenomena. We begin with the application of a reduced-version of the model, to simulate its basic behavior, and follow with the analytical application of the full framework to account for a range of perceptual and cognitive phenomena, including some intractable problems in the philosophy of mind. All the simulations were prepared using in-house code implemented in Matlab (Mathworks), both for the reduced-model and for generating 3-dimensional projective spaces and 2-dimensional geometrical images for illustration of the broader analysis. We used linear algebra and Matlab's opengl rendering engine for patches based on wireframe meshes, either created by us or freely available online (e.g., the human model).

## 3.1  Reduced-computational model

We derived and implemented a simplified version of our model, which included an embodied agent embedded in a "flat land" (2-dimensional plane), with perspective taking and action governed by free energy minimization. This computational model functions as a basic kernel onto which a full projective geometry engine operating on a 3-dimensional virtual world (filled with meshes of objects, avatars and landscapes labeled with personal relevance) will be added in the future. We apply this simplified model because it preserves essential features of the full model and offers a general demonstration of its behavior and utility.

It is a reduced model because it is not based on an optimization of 4*4 matrices of projective transformations, for a fully 3D world. It operates on a simplified model of perspective taking and projective geometry, adopted for the simulations, in a 2D world. The computation of free energy is also simplified to that of surprise per se. (The Matlab code implementing these simulations may be requested to the corresponding author).

### 3.1.1 Definition of the model

The world is modeled as a 2-dimensional finite plane ($nx \times ny \times k$ matrix; $k$ standing for the number of dimensions of personal relevance being encoded as conditional probabilities)(see figure 2a). There is an 'objective' world map $F$, corresponding to actual expected facts, as sampled by the agent's sensory interfaces, and a 'subjective' world-map, $H$ corresponding to prior beliefs or hypotheses about facts stored in memory. Each 2-dimensional location on the maps, with coordinate $X$, is associated with two conditional probabilities, related to personal relevance, encoding safety and pleasure as $P(\text{Safety}|H,F)$ and $P(\text{Pleasure}|H,F)$. These probabilities represent valenced situations or events that are more or less likely to happen at a given location and within a given portion of space (e.g., an object, a room), under specific perspectives according to $F$ or $H$: 'safe' ($S$), and 'pleasant' ($P$). The two probabilities are independent, as we consider that the same locations or portions of space can be the support of events with various degrees of $S$ and $P$. When the probability of $S$ is low situations are likely to be dangerous, when that of $P$ is, situations are likely to be unpleasant. (We note that $F$ and $H$ represent a concrete simplification and unpacking of what we denote generically with $X$ above).

An agent is embodied in the world plane (green dot in figure 2a). Its field of consciousness ($FoC$) is defined as a local compartment of the plane (see box with dashed cross on the charts) with the agent at the center: this is the world that it can perceive and imagine just around it. A subset of the $FoC$ is updated based on sensory evidence $D$ across a possible field of sensation ($FoS$), given the actual orientation of the agent in the world map. The complete $FoC$ at each instant is computed as a linear combination (1) of perceptual evidence, as sampled by the 'current field of sensation' (FoS) over the real-world personal relevance map, and (2) of the spatial complement of the $FoS$ in the $FoC$, as filled based on the current priors in memory at the corresponding locations $x$. The $FoC$ is thus a spatial combination of perceptual evidence and prior beliefs.

**figure 2. Reduced-model simulation results.** A. The results are organized as a table. Each row corresponds to a time period in chronological order from top to bottom. Processing modes alternate between Perception and Imagination (pale blue frame). Real-world personal relevance is represented by a matrix $F$ and the current prior beliefs in memory by a matrix $H$. Personal relevance is color coded from dark blue for "dangerous" (unsafe and unpleasant), through lighter blue for safe, to pink for pleasant situations. The agent can be located at $x$ within these maps. The field of consciousness ($FoC$) encompasses a subset of the world map centered at $x$. It is a linear combination of $F$ as sampled by the field of sensation ($FoS$) and $H$. An intentional vector $T$ (green arrow) points in the direction of maximum safety and pleasantness according to the minimization of free energy (FE). B. Left chart: average FE for the whole $FoC$ in red, and the optimal perspective in green. Right: absolute value of a smoothed version of the first derivate of FE, $d(\text{FE})/FE$, scaled by the current value of FE, as a function of time (number of iterations). C. Analytics characterizing internal and external parameters of the agent (see text). [Model parameters: $n : 400$ (world size, $n \times n$ matrix); $nfoc : 50$ (half span in pixel of the FoC); $nit : 80$ (lifespan: total number of iteration); $wEvP : 0.9$ (Perception mode: relative weight of local sensory evidence over prior beliefs in FoS); $wEvI : 0.3$ (Imagination mode: relative weight of local sensory evidence over prior beliefs in FoS for local environment tracking); $dEth_l : 0.01$ (Threshold: lower Value, for the FE derivative statistics controlling the switch between Perception and Imagination); $dEth_h : 0.06$ (Threshold: higher Value, for the FE derivative statistics. The Imagination mode is triggered when the dFE statistic cross the lower threshold and stops when it crosses the upper one.); $wlocFoC : 0.4$ (Imagination mode: relative weight of local overall FoC in the definition of the intentional vector); $lr : 0.5$ ('Learning rate', number of time perceptual evidence weights in the average $H(FoC, t+1) = [lr * FoC(F, H, x, t) + FoC(H, x, t)]/(lr+1))$; $speed : 7$ (Action: agent nominal speed in number of pixels in the matrix).]. D. Results of simulations for the agent with an impaired capacity for imagination (see text) [Model parameters: $n : 400$; $nfoc : 50$; $nit : 80$; $wEvP : 0.9$; $wEvI : 0.3$; $dEth_l : 0.001$; $dEth_h : 0.06$; $wlocFoC : 0.4$; $lr : 0.5$; $speed : 7$].

The agent constantly tries to optimize its field of consciousness with respect to personal relevance. It continuously scans its $FoC$ with an attentional beam governed by an 'intentional vector' $T$ (see figure 2a) over a scope corresponding to a possible field of sensation ($FoS$) and direction of action. The attentional beam is down-weighted as a function of distance to emulate perspective effects (perspective division). The attentional beam can vary in orientation and field of view. The optimization of the parameters of orientation and field of view is based on the minimization of average free energy (FE) within the FoC across different perspectives (figure 2b). Free energy is simply approximated in this reduced-model as the negative logarithm of the conditional probabilities (or surprise) encoding personal relevance. Intentionality is optimized within a $FoC$ centered at $x$ by selecting the intentional vector, i.e., a perspective $T$, which minimizes FE $= -log(P(\text{Relevance}(\text{FoC(x)})|H, F, T))$. More specifically, the intentional vector is controlled by the direction and field of view for which the average free energy is minimal in the $FoC$ at the current location $x$. It corresponds to a locally optimal perspective $T$ (green vectors on the Figs). As a result (figure 2b) the selected perspective has the lowest average FE (green lines in the figure)(compare with average $FE$ in the overall $FoC$, red lines in figure 2b). The relative contribution of $H$ and $F$ in the computation depends on whether the agent operates in a 'Perception-mode' (high-load of sensory information) or in 'Imagination-mode' (high-load of prior information from memory) (see figure 2a-b). FE is related to motivational drives in the following way. First, in order to maximize the likelihood of pleasure, the agent's intentional vector will tend to point towards directions of anticipated maximal pleasure. This implements a mechanism of 'approach' [31]. Second, in order to minimize exposure to danger, the vector will tend to point towards directions of anticipated maximal safety. This implements a mechanism of 'avoidance'. The intentional vector is the sum of the two corresponding vectors.

Action is governed by the intentional vector, and position $x$ is updated at each cycle, by integrating a simple linear differential equation, based on a unit intentional vector $u$ and a coefficient setting the speed $s$: $x(t+1) = x(t) + su$. Priors are updated at each cycle, based on the current $FoC$. The prior map $H$ is locally averaged with the current $FoC$ in order to reflect the impact of perceptual evidence and conscious experience on beliefs, with a learning rate $r$ that controls the weight of the current $FoC$ in updating past beliefs in $H$.

Following the principles defined above, when free energy becomes too stable (first-derivative stability analysis), for instance when the agent is not finding any location around its current location that clearly improves its situation, i.e., further minimizes free energy, the agent enters into a 'pure Imagination mode' (a form of active mind-wandering), during which it scans its memory (its current prior map) in order to identify locations $X'$ beyond its current $FoC$, at which free energy could be further of more efficiently minimized. The switch between cycles of perception-action and imagination is defined by upper and lower thresholds bounding a smoothed version of the normalised derivative of $FE$: $d(FE)/FE$. The imagination mode is triggered when the lower threshold is crossed and turned off when the upper threshold is crossed. Imagination consists of the access to priors in memory via the generation of a $FoC$ which is based on the prior map $H$ versus fact map $F$, according to a set relative weight (even during imagination, consciousness normally keeps track of its local environment). The imaginary $FoC$ map that minimizes FE is selected (model selection) as a desired destination, and weights in the definition of the intentional vector.

During the Imagination mode, imagination tends to drive action. The intentional vector is defined based on the most desirable location $x'$ provided through a process of active "mind-wandering". The same equation that governs the Perception mode is integrated to actualize the motion, with a set speed parameter.

### 3.1.2 Simulation results

figure 2 illustrates the simulation results. At $T_1$ (figure 2a), the agent starts stuck in an environment which appears worse than expected as it updates its prior map, and wanders around in circles without finding a more optimal outcome from the standpoint of its local $FoC$. Free

energy is high (figure 2b), and stable. The green arrow on the left chart of figure 2b indicates the difference in average free energy content between the selected local optimal perspective $T$ and the overall $FoC$. In other words, the agent has no "hope" in the local area it has been sampling through its $FoC$. At $T_2$, the agent starts imagining itself at a location at which FE is minimal (red arrow) according to its prior map in memory, and thus which maximizes the likelihood of pleasantness and safety. The agent proceeds to move towards this location. At $T_3$, the agent is shown arriving at a location that does not exactly match its prior beliefs about personal relevance. According to its current $FoC$, the location is rather safe but not particularly pleasant. Free energy is lower than before but remains relatively high accordingly (see plateau between the two bursts of imagination indicated by red arrows in figure 2b). The situation is better but far from optimal. As the agent explores the local area, it updates its prior beliefs to better reflect reality ($F$) based on the sampling of the world-map via its current $FoS$. At $T_4$, the agent finds itself stuck in the safe but anhedonic area. The lack of change in FE, irrespective of the perspective being taken, triggers again the agent's imagination mode. The optimal imaginary $FoC$ is now found by imaginary projection in a location with high probability of safety and pleasantness. FE relaxes and drops by anticipation (red arrow), offering a new 'hope' for the agent's $FoC$ optimization. The agent starts its journey towards this new destination. The imagined location happens to fall outside the actual location of the area of pleasantness on the world map (see $F$ versus $H$ maps in figure 2a). But it happens to be near it. When the agent arrives at the projected location ($T_5$), the actual area of maximum likelihood of pleasantness and safety begins to enter in its $FoC$ through its $FoS$, and, in the process of minimizing free energy, the agent moves in this direction. At $T_6$, the agent has now found a broadly optimal solution to the problem of minimization of FE, and remains stuck in it, enjoying its own $FoC$, without further optimization in perspective. FE is low and there is no local path to an even lower FE. After a while, the agent which is 'bored' (the first derivativebecomes stable again FE), starts to imagine other options. But because it derives its non-local inferences entirely based on its prior map $H$ during Imagination-mode, and because that current prior map does not show any better areas outside the current area covered by its $FoC$, the agent goes back and forth locally, oscillating between Perception-mode and Imagination-mode, with its observations globally matching its expectations: it is globally satisfied. Its world-model in memory is now much closer to the actual world-map, but it still carries biases, which make the agent blind to potentially better areas and, as a result, unmotivated to further travel to remote locations, in the absence of any internal or external changes.

figure 2c, shows a set of analytics from the agent. The middle chart displays (colored lines) cumulative average frequencies of different categories of perceived (continuous lines) versus imagined (dashed lines) situations in the $FoC$ as a function of time, corresponding to the proportion of time spent experiencing: pleasantness (red lines), safety (blue lines), and overall utility (green lines), which were calculated as the average of the cumulative frequencies of pleasantness and safety. Overall, the agent optimizes utility: starting from a bad spot, it first experiences a decreasing sense of safety, and as it imagines a better non-local outcome experiences by anticipation an increasing amount of safety and pleasantness. Imagination as a projective solution to the problem of local minima helps the agent to climb the gradient of utility, and thus renders it more resilient. Because the agent learns by progressively updating its prior beliefs, its representation of the world in memory also progressively becomes more accurate as it explores the world (black line in the middle chart, which shows the correlation coefficient or predictive power between the prior map $H$ and the world fact map $F$). The top chart shows the difference in cumulative frequency for overall utility between perception and imagination or the agent's level of "optimism". The agent starts with a phase of disillusionment, followed by a phase of optimism, progressively converging toward a balanced state of realism (the expectations of the agent better reflect the world). The bottom chart shows in color (from dark blue for low values, through green, to dark red for high values), the cumulative frequency of three parameters related to covert and overt activity: (1) the angle of the attentional beam, (2) the quantity of rotation of the intentional vector, and (3) the amount of actual displacement. The agent's overall activity decreases as it

optimizes its $FoC$ climbing the gradient of personal relevance, and its attentional beam tends to become narrower. figure 2d shows the results of another simulation with identical initial conditions, the same agent, except for a lower value of the threshold controlling the triggering of the imagination mode as a function of the dynamics of the first derivative of FE. The agent has a dramatically impaired ability to rely upon non-local projective imagination, e.g., third-person perspective, and only experiences the world through a local $FoC$, e.g., first-person perspective. As a result, its exploration of the world is much more limited (see maps at $T6$). FE minimization follows a slower and smoother optimization. The level of FE remains high and sub-optimal at the end of the agent's journey, where it finds safety but not pleasantness, by mere proximity. The charts showing the agent's analytics (next to the corresponding chart for the first agent in figure 2) indicate bad performance in optimization of utility, with amelioration in experienced safety, but barely any pleasantness experienced. The accuracy of the agent's representation of the world is also lower. The agent displays little optimism. The lack of imagination has rendered the agent less resilient.

These simulation results demonstrate the suitability and functioning of the general mathematical instantiation of the PC Model. We now concentrate on applications to more specific psychological phenomena, which we explain by reasoning based on the full model.

## 3.2   The puzzle of the missing origin

According to the phenomenological postulates, consciousness seems to embed a vantage point that is elusively connected to the center of lived space and plays a generative role in shaping lived space. The elusiveness of this vantage point is predicted by the PCM model. It follows from an integral, definitional property of a projective space. colorred We saw that a definition of the $n$ dimensional projective space $\mathbb{P}^n$ can be obtained from a linear space $\widetilde{V}$ of dimension $n+1$, by considering all the straight lines from the origin $O$. This can be understood as an embedding of the space $\mathbb{P}^n$ in the space $\mathbb{P}^{n+1}$, with an additional general point of view in the supplementary dimension. The central point from which the projection is set up must be excluded from the projection to generate a projective space. By analogy let us consider the case of a projection on a 2-dimensional plane as in perspective painting or photography: to project 3-dimensional objects in Euclidean ambient space onto a canvas or film so that they appear in perspective, a center of projection or focal point has to be chosen outside of the plane of projection in ambient 3-dimensional space, otherwise all the objects would be projected onto the same point on the 2-dimensional plane that serves for representation. In other words, 2-dimensional images are really understood as 3-dimensional objects, with an elusive origin residing in a supplementary, implicit third dimension, which is necessary to set up a projection. This property generalizes to projective spaces of higher dimension such as 3-dimensional projective spaces like our lived space: to project 3-dimensional objects from a certain projective space (outside or in memory) in our projective space of 3 dimensions, a center of projection or focal point has to be chosen outside that space in a 4-dimensional space, which in practice requires the use of homogeneous coordinates in order to make the points in 3-dimensional ambient space correspond to 4-dimensional objects. Note the important theorem [110] asserting that the full projective group of a projective space of dimension $n$ can be obtained by composing perspectives in a projective space of dimension $(n+1)$. Consequently, the embedding of our $3D$ space in the $4D$ space allows us to generate projective changes of frames, as they are used in imagination, using perspectives from imagined points outside of the world.

In other words, the 3-dimensional lived space can be embedded in a 4-dimensional projective setup, with an elusive origin residing in a supplementary, implicit 4th (vector) dimension, allowing the projective group to act as changes of points of view.

Nevertheless, this extra-dimension is not mysterious. It is a parametric dimension, which controls perspective taking by configuring the direction, scope and perspective of the projection. In the numerical model, it is expressed in the use of $4 \times 4$ transformation matrices $T$ (applied on homogeneous coordinates). The extra dimension with the excluded origin is essential in defining and generating the projection. It is fully computable but it does not have a direct geometrical (or

subjective) manifestation in the 3-dimensional space. In 3-dimensional lived space, this extra dimension remains inaccessible. This means that according to our model, imagination as applied to the phenomenological 3-dimensional world, is generated from elements of a perception spectrum residing in a higher-dimension that cannot be directly grasped through a 3-dimensional representation of space.

## 3.3  Bistable illusions from ambiguous planar figures

The PCM can account for bistable visual illusions such as the ubiquitous Necker cube. Planar projection of objects in spaces or "images" tend to be decoded by the brain as representations of 3-dimensional objects from a given standpoint and interpreted as such in perception [111]. In other words, in such contexts, using our model, somehow perception always engages a situated form of imagination: it extrudes depth information as it renders images in consciousness, but it remains driven by sensory data so it appears as a perception. The Necker cube illusion consists of the 3-dimensional illusion of a cube switching orientation, which is elicited by a visual confrontation with a 2-dimensional ambiguous figure (see figure 3-a). The ambiguous figure can be interpreted as the planar orthographic (affine Euclidean) projection $I$ of the edges of a generic cube $R$ (serving as model prototype in memory) on an image plane. The projection is ambiguous: two different orientations, $R_1$ and $R_2$, of the cube in ambient space can have identical orthographic projections $I_1$ and $I_2$ on the image plane. In the experiment, the projections on the retinas of such ambiguous image plane provide sensory inputs $D$ to the brain.

Based on the PCM, such an ambiguity would challenge any attempt by the active inference engine to solve the inverse problem of ecologically interpreting sensory inputs $D$ as a likely generic cube $R$ in 3-dimensional space with a given orientation.

Studies of the illusion have emphasized local rules for controlling the switch of the Necker cube (e.g., Hoffman's rule), and have considered the problem in relation to the "generic viewpoint assumption", which implies that there is no a priori viewpoint position defined in such images but that a viewpoint is nevertheless assumed in interpreting it [112]. In this framework, the bistability of the illusion induced by the ambiguity has received a Bayesian formulation [112]. Our model complements this approach by specifying the projective geometrical underpinning of the phenomenon.

Solving the underlying inverse problem involves perspective taking: a projective transformation $T$ that best fits the implied orientation must be selected to represent the cube in projective space. Unless strong priors are given about the object being of type cube $R$ from the outset, in general, the selection of $R$ as a cube is also part of the process of inference. In our framework, this corresponds to maximizing the posterior probability $P(T, R|D) = \frac{P(D|T,R)P(T,R)}{P(D)}$ of the transformation $T$ and of $R$, as a generic cube and internal model in memory, given $D$ from sensory information. If for an optimal selection of $R$ the maximum of P(T,R—D) was unique, i.e., there would be a unique $T$ maximizing the quantity, and conscious access could be computed in a stable manner as (the prediction) $S = T \circ R$, which is the generic equation of the rendering of a cube in perspective with a genuine sense of depth and a given orientation in a projective space.

In the normal perception of a real cube in ambient space, extraction of color and shade information, but also perspective effects and binocular disparity related to depth, would constrain the inverse problem. The two viewpoints would not have an equivalent posterior probability (i.e., there would be a unique optimal solution to the problem of selecting $T$) . With a real wireframe cube presented at finite distance and projected on a 2-dimensional image plane, perspective and binocular disparity effects would remain to constrain the inference (e.g., the front side would appear bigger than the back side on retinal images, and there would be different angles of projection) (figure 3-a, left-tier). With the Necker cube's implicit orthographic projection on a 2-dimensional image plane, there are no perspective cues, and there is no binocular disparity (figure 3-a, right-tier). The transformations $T_1$ and $T_2$ that correspond to $R_1$ and $R_2$ respectively, i.e., the cube under the different possible viewpoints, have the same posterior probability and thus

the same level of free energy (FE).

**figure 3.   Account of psychological phenomena.** A. **projective solutions to the Necker cube.** Projection of two complementary views $S_1$ and $S_2$ of a generic wireframe cube $R$ in a projective space (rendered on an image plane). **Left**: The point of view is located near the projected cube in ambient space: perspective information helps to disambiguate the inference about the two possible orientations of the cube, $T_1$ and $T_2$ (see gray ellipses indicating the front). As a result, free energy ($FE$) always possesses a minimum for a unique transformation $T$ and perception can stabilize. **Right**: The point of view is placed far away but seen with a strong zoom. No perspective information remains to disambiguate the orientation of the cube. Free energy does not possess a unique minimum, each option is unsatisfying, the inference is globally undecidable but irrepressible, and the system can only oscillate between the two possible outcomes. B. **Projective solution to VR-mediated dissociations in phenomenal selfhood**. The visual field is indicated (dashed-green line) as a subset of lived space. **Left**: Projection from the normal perception spectrum: $S_1 = T^P \circ R$, the projective frame is calibrated in Perception-mode. **Right**: Projection from the pure imagination subgroup: $S_2 = T^I \circ R$. Because of the experimental conditions, a projective solution that is not normally used for perception will be chosen as the best explanation of sensory data. C. **The anti-space beyond the plane at infinity**. Rendering of the projective space with a transparent plane at infinity and no clipping, revealing the involution of the space mirroring the back of the ambient space (see text). D. **The ”God's eye” vantage point on the projective space**. Rendering on a 2-dimensional image-plane of the projective space and its mapping of the world model (used in (b) and (c)), from the ”God's eye” point of view at infinity. Almost all the space is visible but completely warped and the world model manifests as structures with complex symmetries.

Moreover, in the group of projective transformations, the only transformations that are compatible with the absence of both perspective cues and binocular disparity entail an implicit point of view placed at a very far distance from the object, near projective infinity $T(\infty)$, with a small viewing angle so that the far distant object (the cube) appears big enough in the scope of the projection. For instance, observing very far objects with powerful binoculars results in their perspective cues vanishing and there is barely any binocular disparity. On the contrary, observing objects from up close increases perspective cues and binocular disparity. In perspective painting, the implicit point of view (center of projection) is generally close to the main group of objects, and the wide angles used for close-ups in photography magnify perspective effects. Points of view placing objects near projective infinity (or equivalently points of view placed near projective infinity) correspond to a subspace of the possible configurations of a projective space, which is equivalent to a space of orthographic projection. Near projective infinity, the 3-dimensional projections of the two alternate orientations of the cube ”look” identical and thus remain totally ambiguous. There is nothing in conscious access allowing the agent to further disambiguate the information and privilege one orientation over the other.

However, placing the point of view near an infinite distance from the object in order to explain the absence of perspective cues and depth disparity in $D$ directly conflicts with ongoing information about the actual distance of the object in ambient space in the experiment, e.g., just in front on a computer screen.

By the end, (i) with no unique maximum in the posterior probability of competing transformations $T$, and (ii) conflicting priors about relative distances, the inverse inference is maximally ambiguous and, outside of a metacognitively imposed voluntary control exerted on higher-order attention, the active inference engine can only oscillate between bistable percepts, with no help from conscious access to stabilize the representation.

## 3.4 Proprioceptive illusions and out-of-body experiences

Clinical and experimental situations can elicit experiential dissociations between the sense of the location of the subjective point of view and the perceived location of the body in ambient space, e.g., OBEs, syndromes of somatoparaphrenia, anarchic hands and phantom limbs, the rubber hand illusion, enfacement or full-body illusions [14, 55]. In OBEs, subjects hallucinate their own body from a third-person perspective, e.g., as a phantom body above them, or travelling away [113].

VR-based manipulations have facilitated reliable elicitation of full body swap illusions in the laboratory, which are analogous to OBEs [14, 47, 55]. For instance, when a participant and a humanoid virtual avatar are being stroked simultaneously, an embodied illusion of spatial identification with the virtual avatar, for instance projected in front of the actual participant, is experienced. The location of the body feels both near the point of view and at the avatar's location. Likewise, Ehrsson [114] placed cameras behind the backs of participants and projected the image of the body of the participants in front using a VR device. This sufficed to induce a subjective impression in the participants of being literally seated behind their own backs and looking at themselves from this location [115–117].

Such dissociations between the subjective locations of the point of view and of the body in space are predicted and can be explained by our model. Base on the PCM, the body as experienced is accessed in consciousness in a projective way, like a multisensory-informed virtual body avatar. In normal perception, the projective body is rendered from an egocentric, first-person vantage point, as an object in space, which serves as a central reference to which the implicit point of view is attached. Only a subset of generalized, degenerate projective transformations $T^p$, the perception spectrum, will be assumed as priors for spatial inference and perspective taking in order to access the body representation projectively as $S(B) = T^p \circ B$.

In VR paradigms inducing the perceptual experience of sitting behind one's own back [114] or similar phenomena of body swapping [47], multimodal sensory data $D$ are conflicting (e.g., visual and somatosensory) in relation to the perspective normally assumed in perception. They are evidence for models of perspective taking, i.e., projective transformations $T^I$, that normally are only associated with pure imagination, e.g., in which the representation of the body is accessed from a third-person perspective. These transformations by no means require an actual displacement of the person in real space: they are a parameterization of the transformation matrices through which the representation of the body is accessed in consciousness, through perspective taking (see figure 3-b).

In such VR paradigms, the process of spatial inference is challenged. Perceptual transformations $T^P$ are expected, i.e., the prior $P(T^P)$ is high, because the process remains driven by sensory data $D$. But the conditional probability of the data $P(D|T, R)$ is maximal for a specific transformation $T^I$ that belongs to the full group of invertible projective transformations, which is normally dedicated to imagination in our model. In such ecologically abnormal situations, the best bet for the agent is to use transformations $T^I$ to explain sensory data $D$. The predictive power of the solution outweighs the priors of normal perceptual transformations. What would normally be only a possibility for the imagination thus manifests as perception, since the inference remains driven by sensory data $D$. The boundary between imagination-related and perception-related modes of conscious access is shifted in this highly unusual context.

## 3.5 Self-representation and pre-reflective self-consciousness

### 3.5.1 Self-representation as awareness of the workspace itself

A model of consciousness should be able to account for the central yet puzzling capacity of consciousness to somehow be aware of itself. This capacity for self-representation or self-acquaintance is an intense topic of debate in the philosophy of mind and cognitive science [?, 5, 51, 57, 118].

Maintaining a certain neutrality about many contested issues here (e.g., the exact nature of the

representation or acquaintance involved), we approach the problem starting from our definition of consciousness as a workspace with a characteristic spatial phenomenology based on projective geometry. In this context, consciousness represents world structures and itself in a spatial manner. Full self-representation for that model of consciousness could mean the ability to represent the entire space of representation itself, including the point of view.

When the point of view is placed near infinite distance from its target and the field of view is maximal, much of the space appears wrapped into the scope of the projection (see below). The entirety of the generative space cannot however be fully encompassed in this scope, since the origin of the space and the parametric extra dimension will always remain elusive.

Conversely, when the point of view is taken from somewhere near us, a kind of imaginary cyclopean eye, the world is roughly divided in two parts: the world at finite distance (itself divided in body space, prehension space and near locomotion space, [11]), and the infinitely far world, where we can only imagine things happening, but out of the scope of our own actions. Nevertheless, consciously, we can relate these things that are projected on the plane at infinity. This ability is consistent with the existence of natural projective metrics, in which all distances are finite, and in which the plane at infinity corresponds to the set of points at maximal distance from the center of vision.

### 3.5.2 The anti-space and Pre-reflective self-awareness

Nevertheless, projective spaces embed further features of genuine self-representation that are remarkable. The mechanism of perspective taking defines a vantage point and a plane at infinity, and the complement of these elements (the represented world) in the generated space has the structure of a 3-dimensional affine space (on which a Euclidean metric can be defined). This geometry is non-trivial. Considering a large sphere, with a given fixed center (the standpoint) and a very large radius that tends to infinity (the possible horizon of consciousness), antipodal points on this sphere project to the same point at projective infinity in the projective space: the space is folded onto itself through the topological gluing of antipodal points. Thus, paradoxically, the back of the ambient or modeled space with respect to the vantage point (be it empty or informed with a representation of a world $R$ from memory), is actually projected in front, but on the other side of the infinity plane, similar to a 3-dimensional virtual mirror image. This "antispace" is usually clipped in digital imaging.

Such strange projective property of involution of the space could not manifest in normal perception because the projective space is practically clipped from residual projections in the compartment beyond the horizon: in perception, the celestial sphere saturates the horizon with perceptual information, and world boundaries in the finite compartment of the space (e.g., walls) will tend to mask it. Moreover, one can expect that for computational efficiency, the parts of the world representation $R$ mapped in memory that are behind the vantage point and outside the scope of the perspective do not generally need to be processed for normal conscious access. They can be ignored by attentional selection, prior to projective inference.

Nevertheless, such a symmetric mapping does exist and is always defined in projective spaces no matter which structures from $R$ fill it, and it is mirrored in the very scope of the projection beyond the plane at infinity, and could thus perhaps be explicitly observed if the horizon could be somewhat rendered less opaque. Thus, geometrically, unbeknownst to us, in the projective space of our consciousness, in perception, imagination and action programming, one also always aims at one's own back, at an "anti-space", when one aims frontward at the horizon (see figure 3-c).

In our framework, this can be interpreted as one way in which consciousness is always tacitly aware of itself prior to any explicit reflection. Geometrically, the vanishing point I am aiming at, the horizon of my conscious space at positive infinity in front, is also the projection of the back of the antipodal point that is located behind me at negative infinity. My implicit vantage point is paradoxically located "between" the same unique ideal point at infinity, which seem at the same time to be two points along opposite directions. Although these statements may appear contradictory they are mathematically well defined in projective geometry, consistent and fully

computable, and the behavior of the space in the finite compartment respects usual relations of orientation and incidence. Thus, according to the model, there is a projective machinery that is defined but remains normally hidden. The way it works as a projection does not imply that an explicit curvature will be apparent in such "elliptical" spaces.

In the variegated and multi-disciplinary literature on self-representation, the concept of "pre-reflective" self-consciousness has become an important subtopic of debate [4, 5, 119–121]. The idea that consciousness always involves a form of pre-reflective self-consciousness is the hypothesis that there is a genuine reflexive (versus reflective) form of self-consciousness that (i) does not require a distinct, higher-order mapping or meta-representation $S_2$ (a representation of a representation) of subjective consciousness $S_1$ at the first-order, (ii) is truly embedded in the space of awareness itself, as elusive as it may be, and (iii) is a necessary basis for higher-order, reflective self-awareness. Interpreted variously, the notion has been criticized for being unanalyzable, ill-defined, or circular, whereas some, on the contrary, emphasize the importance of finding models that could incorporate such circularity without generating paradoxes [4, 5].

We suggest that the peculiar topological properties of projective geometry could perhaps make sense of the hypothesis of a pre-reflective consciousness, in a more formal way. This is at least one way in which a pre-reflective form of self-awareness can be built into the architecture of consciousness, and it is an intrinsic feature of our model.

### 3.5.3 Gaze behavior in mind wandering

Beyond its general role in shaping imagination under the driving of personal relevance, this peculiar geometry might also further illuminate the phenomenon of mind wandering. Mind wandering and daydreaming are directly linked to increased self-referential processing and self-consciousness [66]. When our mind wanders it tends to turn toward self-related inner concerns away from external information processing [122] and to generate projective scenarios about the past, the present, and the future [66]. Attention is withdrawn from outside events, we are less sensitive to sensory data, and we may appear to others as phasing out or being distant. Our eyes seem to look through the vacuum to the external observer; that is, we lose binocular fixation on the proximal environment, and eye movements can become erratic [123]. In other words, we seem to weight down the processing of multisensory data $D$ and weight up that of information from memory and internal simulations on the $R$ being accessed, increasing, according to our model, imagination (about past, present and future events) over ongoing perception. Interpreted from a geometrical standpoint, the overt gaze behavior may suggest that we tend to enter into divergent parallel binocular states during self-referential processing. Interpreted as an index of a projective parameterization of a perspective on an imaginary space, divergent parallel binocular states imply underlying projective transformations $T$ that aim at or from infinity. Although this remains speculative and should be addressed experimentally in the future, our model would predict that, in such aiming, the subject may also be aiming at the anti-space, i.e., at her or his own back. Under such perspectives, a subjective perspective on the entire space might be approximated, in a manner that is consistent with an activity of self-representation.

### 3.5.4 Mystical and psychedelic experiences

Various spiritual practices (e.g., meditation), the use of hallucinogenic drugs (e.g., mescaline), extreme life experiences (e.g., near death experiences), dreams, phenomena of entoptic vision induced by ocular pressure, and clinical conditions have been associated with reported subjective experiences of strange and fascinating spatial patterns in consciousness: radial patterns, spirals, funnels, cobwebs, honeycomb lattices and triangles, deep tunnels, which have then been a topic of artistic representation in many cultures (e.g., mandalas) [124–128]. Such patterns have been hypothesized to reflect the inner structure of spatial consciousness as a level of representation and a mode of access to information [126], as cited in [127]. We can wonder whether our model can

predict similar patterns, as would be expected from a valid model of consciousness, given their importance in human experience.

Projective spaces and transformations can lead to peculiar perspectives on the structure of the space, in particular when the implicit standpoint is placed at near infinite distance from the center of the world in the scope of the projection (which is a parameter to set in the corresponding transformation matrix). This is, so to speak, the "God's eye" point of view. From this standpoint almost the entire space of representation manifests in the scope of the projection. As a result, symmetries and strange topological patterns may appear, which are reminiscent of the type of aforementioned patterns (see figure 3-d). We hypothesize that such psychological phenomena are a reflection of the application of "limit" projective transformations (e.g., placing the point of view near infinity), generally in a context of extreme conditions, in which the mind moves away from the normal constraints of perception and uses the imagination group widely to solve the inverse problem of representing a spatial world in a projective manner from a stream of sensory and mnemonic information. We may wonder whether such a projective setup, with its intrinsic dissociation between the subjective standpoint and the apparent location of the body in space, could be a natural ground for the emergence of beliefs in various metaphysical forms of dualism [129].

# 4    Discussion

We derived a mathematical model of embodied consciousness, the Projective Consciousness Model (PCM), based on projective geometry and active inference, which features a characteristic mode of representation of space or spatial phenomenology, resulting from a mechanism of conscious access based on perspective taking, and which operates based on the minimization of free energy as related to a probabilistic field encoding cognitive and affective personal relevancies mapped across space. The model: (i) yields basic behaviors that are predicted by the theory and phenomenologically sound, and accounts for a variety of cognitive phenomena that have not been linked previously in a unifying manner, (ii) predicts core aspects of the characteristic phenomenology itself, and (iii) renders explicit the link between phenomenological features and their functional and computational value as a specific mode of access to information and a support for the driving of action in the context of situated cognition.

The PCM directly connects consciousness, active inference and appraisal to resilience, in line with previous hypotheses [1, 39]. So far no model had been introduced, which was both fully computationally tractable, and integrated a comprehensive model of cognition and affective processing in embodied ecosystems, including a model of the first-person or experiential perspective. The development and implementation of mechanistic models of cognitive and emotional resilience are essential in order to precisely understand, in a predictive manner, the mechanisms mediating resilience, in the perspective of preventive and acute healthcare.

## 4.1    Neurocomputational hypotheses

The computational architecture implied by our model can be translated into general neurocomputational hypotheses, in accordance with the principles of neurophenomenology [8, 130]. We have argued that such and integrative processes of conscious representation and control may rely on a form of virtualization in the brain, integrating information processing through layers of abstraction into a global, resilient predictive world model [1].

In this framework, we expect the relationships between the PCM algorithm and the functioning of neural networks, in particular at the microscale, to be complex and indirect. On a general note, assuming that the brain uses a projective geometry as part of its generative model is agnostic with regard to the particular neuronal encoding. For example, positions and metrics could be encoded as continuous variables in neuronal population activity. Conversely, there could be a discrete state space model of the sort suggested by classical receptive fields (with compact support). Irrespective

of the particular encoding, our point is that there must be a formal connectivity among neuronal populations encoding allowable transformations that mediate belief updating among neuronal representations (see for instance recent hypotheses about neural computation referring to topological theorems that are also intrinsic to projective geometry [131]).

In terms of the functional architectures that the projective geometry model calls for, there is an abundance of empirical evidence implicating specific brain systems. On this basis, the model predicts that the brain should embed two main engines that are hierarchically coupled (figure 4): (i) an inference engine grouping higher levels of control concerned with homeostasis, emotion, memory, language and executive functions, or more generally personal relevance; and (ii) a lower (sensorimotor) projective geometry engine, relying on multisensory integration and motor programming and concerned with representing the world and the body in the world and mapping gradients of free energy to drive intentionality. Active inference provides the overall hierarchical infrastructure and various message-passing schemes required for generating predictions and inferring the world ($R$), adaptive transformations ($T$).

**figure 4. Macroscale neurocomputational model** Lateral aspect of the right hemisphere of a brain model. The projective geometry engine is implemented in posterior cortical and subcortical systems (gray blue). The framing and rendering of the subjective experience of space through perspective taking is performed by those systems, which apply projective transformations such that $S = T \circ R$. The active inference engine is implemented in anterior limbic and non-limbic cortical and subcortical systems (red). It integrates memory (including world-representations $R$ stored in mesiotemporal systems), affective processing of personal relevance (the amygdala and ventral prefrontal cortex), and executive components (the prefrontal cortex and basal ganglia). The cerebellum (bright blue), along with the basal ganglia, perform an ongoing sensorimotor calibration, relating the projective space to the Euclidean space of action $\mathbb{E}$. All the systems are coupled and interact through large-scale connectivity and networks (color arrows).

From an anatomical standpoint, we hypothesize that the inference engine relies on anterior cortical and subcortical systems, including limbic and non-limbic frontal and temporal association cortices, the amygdala and hippocampus. The projective geometry engine can be hypothesized to rely on posterior temporal-parietal-occipital, modal and multimodal sensory systems, in particular in parietal systems, encompassing exteroceptive, proprioceptive (including vestibular) and interoceptive processing, but also on frontal premotor regions. Recent evidence supports the existence of a partition of the cerebral cortex into two main anterior and posterior functional ensembles, in a manner that is compatible with our hypothesis of a posterior real-time projective geometry engine and of a slower anterior hierarchical active inference engine quantifying personal relevance [132].

The tight relationship between spatial memory and affective processing in the brain, as mediated by the interactions between the hippocampus and amygdala, is in general support of a model of space in memory imbued with personal relevance: the strong interactions between the hippocampus and the amygdala and other limbic structures are thought to play a role in the labeling of memory of spatial locations with personal relevance [31, 37]. A large body of literature implicates the hippocampus in spatial representations and scene construction; ranging from place cells through to sequential navigations through metric spaces [133–135]. The para-hippocampal system would then be expected (in our model) to play a key role in the representation $R$ that is stored in memory and that is accessed in consciousness through projective transformations in the recall of scenes. Likewise, the strong interactions between the (DMN) and medial temporal systems in the context of autobiographical memory recall [136] can be understood in the PCM framework as related to the process of addressing $R$ and its labeling in terms of personal relevance in the model.

The construction and application of projective transformation $T$ to both sensory data $D$ and data from memory $R$ are presumed to call on parietal processing. For example, the distinction between what and where pathways in the brain [137, 138] tells us immediately that a vast amount

of the visual hierarchy (the dorsal stream, which includes much of the parietal cortex) is devoted to spatial transformations. This observation is endorsed by spatial neglect syndromes following damage to the parietal cortex [139, 140]. Furthermore, mental rotation tasks [141] and scene construction tasks [142] almost universally engage parietal cortex, including the precuneus and retrosplenial cortex. Work looking at saccadic eye control (perhaps the most obvious manifestation of spatial transformations) and sensorimotor transformations in motor planning implicates regions like the intraparietal sulcus [143].

At a more global scale, we expect frontal-parietal-temporal interactions to be related to inference and perspective taking via the adaptive selection and application of projective transformations to spatial models in memory as a function of personal relevance. Conscious-access across cycles of inference requires addressing information related to $R$ in memory or to $D$ in multisensory systems in an adaptive and well-scheduled way based on such transformations. Switching between these two perceptual and imaginary modes of access may be expected to be supported by medial and lateral fronto-parietal networks. There is an ongoing debate in the cognitive neuroscience of consciousness as to whether conscious access and experience require fronto-parietal interactions or only coherent activity in posterior cortices [18, 19]. We believe that this debate might be in part driven by a focus on different aspects of consciousness that may end up corresponding to the different hierarchical levels implicit in our model and its division into two main engines.

At a lower level of integration, the brainstem, thalamus, reticular nucleus of the thalamus, the precuneus and related regions can also expected to play a role in the arbitration or switch between different projective modes (e.g., "Perception-mode" versus "Imagination-mode"), and in a form of arousal related to free energy. These brain systems can be expected to modulate, notably through mono-aminergic transmission, the precision or weight allocated to sensory data $D$ in the process of projective inference, and are key modulators of vigilance states. The cerebellum as well as the basal ganglia are expected to play a key role in the calibration of the different subspaces and transformations between them (e.g., between $S$ the projective space of conscious access based on perspective taking, the Euclidean representations of knowledge in memory $R$, which is independent from any point of observation, the associated field of personal relevance, and the compilation of predicted action into a space of effectors with biomechanical constraints, figure 4).

## 4.2   Real-world applications: implications for system design and autonomous computational agents

Beyond the scientific and theoretical value of a computational model of embodied consciousness, our approach has a strong potential for real-world impact. We argue that the emerging set of interrelated fields that include autonomous computational agents, robotics, and artificial intelligence (AI), as well as new developments in virtual reality (VR) applications, can benefit significantly from the development of a mathematical model of embodied consciousness. The PC Model contributes by specifying what parts of consciousness (beyond intrinsic biological properties) could be transferred into an AI environment and the principles by which these components will function and become integrated. The model also provides specific predictions for how embodied AI agents will behave and an explanatory framework by which these predictions can be interpreted in human terms. This latter advantage is no small contribution: it has become increasingly clear that one of the most pressing problems with the application of advanced AI systems is their ultimate inscrutability [144].

Some of the principles at the center of our model are actually already heavily used, in bits and pieces, in industry: in filtering and image processing, in cybernetics and robotics, in cinema (CGI), in virtual and augmented reality. Our model, which can be interpreted as a formal theory of "enaction" and embodiment, though without the externalist implications normally associated with these approaches, see [8], integrates these principles into a coherent whole that offers the sketch of a design for embodied consciousness in artificial systems.

## 4.3 Real-world applications: consciousness and Virtual Reality

If the PCM is correct, consciousness accesses multisensory information $D$ by framing it as a representation of space in a 3-dimensional projective space, supported by prior beliefs about the 3-dimensional structure of the space to be represented and about the forward models governing sensory processing. In this context, the framing of spatial consciousness by projective geometry may explain why it can make sense of 2-dimensional projections such as images and films from multiple standpoints, and why Virtual Reality (VR) can be so effective at inducing an immersive feeling of realistic presence in a 3-dimensional world.

The PCM predicts that configurations of inputs provided to sensory interfaces (i.e., a projection of a geometrical model of a virtual world or of the real world on the sensory interfaces) will be interpreted and framed in consciousness as the experience of a world of 3-dimensional objects in perspective. In VR, the world represented, just as with normal perception, will be based on an inverse projection inferring the causes of the sensory inputs and implicating an operation of perspective taking. If the virtual world is simulated in a realistic manner, this simulated world is expected to appear to consciousness with a sense of direct realism, that is with a sense of natural evidence, similar to that of normal perception.

VR immerses participants in an artificial sensorimotor environment that takes cybernetic control of the interfacing of the conscious workspace with sensory inputs and motor outputs [145]. Through metacognitive processing, we may well understand that cognition is immersed in a counter-factual virtual world, nevertheless the brain interprets the inputs and the consequences of the outputs on them, at a sensorimotor level, in an embodied and affective manner, as it would with the real physical world. Consciousness cannot help but access this information as it would for normal perception, since it is structured to be processed accordingly. The generation of the stimuli used for VR-devices (e.g., Head Mounted Display; HMD), is based on the application of projective geometry (and projective transformations) to 3-dimensional geometrical models of dynamical structures in Euclidean space, in a manner that takes into account the real-time position and orientation of the HMD for perspective taking. Sounds are also presented through headphones according to geometrical and physical transformations to account for the localization of their sources in the world model. A persistent problem with VR experiences for decades has been the ubiquitous feelings of motion sickness and nausea. VR fails to be immersive and to induce a rich, natural sense of presence in another world, when the sensorimotor contingencies, the responsiveness of the VR devices to motion (HMD, audio headset, controllers tracked in space), and the sense of agency and the accurate control of intentionality and action, are limited by the technological setup. In other words, immersive presence is only achieved when the setup can flawlessly simulate the input-output relationships and parameters that condition the type of perspective taking embedded in the PCM.

In turn, because its geometrical properties are so in tune with the generative mechanisms of consciousness we put forth, VR appears as an exceptionally promising tool for studying consciousness. The application of such VR paradigms in the context of paradigms of sensory conflict can reliably induce numerous states of bodily awareness in which phenomenal selfhood parameters can be manipulated at a sensorimotor level independently, from: (i) the self-identification with a body as whole; (ii) the self-location relative to the lived body; and (iii) the framing in first-person versus third-person perspectives [14, 47, 55]. We demonstrated above how the PCM can account for such illusions and phenomenal dissociations, and how it can make sense of their paradoxical emergence in the context of non-ecological VR-mediated experimental manipulations. VR could be further used to test hypotheses about real, normal and pathological human behavior based on direct predictions of the PCM as a function of its parameters, These types of manipulations in VR environments would offer experimental paradigms, based on formal mathematical principles, to quantitatively study processes such as optimal decision-making under stress as well as the structure and dynamics of consciousness.

## 4.4  Limitations

The model proposed here is only a starting point and a very primitive version of what we could expect from a mathematical and computational model of consciousness ultimately, in terms of complexity and formal rigor. However, we see the principles introduced herein as a basis for constraining more comprehensive models of consciousness and for exploring the relationships of consciousness to other scientific theories. Examples of the diverse applications and mathematical structures it implicates vary from topology to group theory, dynamical systems theory and beyond, but also for making connections with clinical and even pedagogical questions.

An essential question that remains open is that of the type of "field" and support the projective space is operating on in consciousness: the very fabric, so to speak, of the conscious space (e.g., a field of numbers?). If projective transformations operate in conscious access and provide us with the rules for the rendering of key spatial aspects of subjective experience in consciousness, the ontology or material support of the interface or medium for this rendering remains mysterious. We suggested that the whole workspace was somehow "virtual", a sort of "user interface", but this does not fully address the question. Images can be rendered and projected on a screen, but what constitutes the 3-dimensional "screen" of consciousness will remain unsettled here. From this standpoint, it is however important to mention the fundamental result established by Hilbert and Artin [87] according to which geometrical axioms of incidence that do not use any kind of numbers in their formulation (like Desargues' and Pappus' axioms) are sufficient to generate by themselves a commutative field of numbers. Numbers appear in an elegant manner through automorphisms of special automorphisms. Therefore, the construction of projective spaces from linear algebra does not lack generality. Moreover, projective geometry can contribute to the emergence of implicit computations, including comparison of ratios of length, like harmonic proportions [110].

It is certain that the structure of the conscious workspace is more complex than what this model captures. There is evidence and a rationale for suggesting that visual awareness, for instance, embeds a hierarchy of structuring spaces, including, at the top of the hierarchy, abstracted spaces representing the Euclidean properties of the world [48]. However, projective geometry is a quite general geometry. When associated with a continuous field with metrical properties, it actually embeds affine and Euclidean geometry, and orthographic projections respecting Euclidean properties can be generated from projective geometry by assuming a vantage point near infinity.

The relationships between the PCM and the dynamical and intensive qualities of consciousness need more elaboration. There are dynamical components in our model, since it is based on active inference, as driven by free energy, which defines the principle and driving force for intentionality and action, and implements a field of motivation based on the spatial mapping of personal relevance. The model features a complex interplay between cycles of perception, imagination, motor programming, and action, through projective transformations and complex mechanisms of inference related to personal relevance. These we expect must fundamentally constrain the pace of time-consciousness, as a moment-to-moment construction of unified experiences involving "protention and retention" [2]. But the exact dynamics that link the inference engine and the projective geometry engine remains largely undefined and underdetermined: it will depend on the level of sophistication and specifications of future implementations of the basic kernel of the model and probably on the species of consciousness targeted. The status of intensive dimensions, e.g., connected to arousal, vigilance and emotions [9], as well as their differentiation into a range of qualitatively different emotional states, need to be worked out, starting from their connection to free energy [92, 93]. Phenomenologically, we proposed that several aspects of bodily feelings are expressed as spatial transformations of the representation of the body in the workspace. To some extent, intensive dimensions, e.g., as with the feeling of more intense cardiac sensations when the heart rate increases [146], can be construed as spatial representations, e.g., as the heartbeat increasingly occupies our attention and interferes with our ability to carry on other mental tasks. In other words, we expect that some intensive dimensions of subjective experience will reduce to spatial quantities that are consciously accessed as such, e.g., the amplitude of a change in posture as a function of the intensity of an emotion, in connection to a field of personal relevance.

Nothing has been said here about many other key dimensions of experience: the gamut of sensory qualities, from colors to tones and smells. Even though again, one might argue that essential aspects of the phenomenology of vision, smell and sound have a spatial manifestation, some key phenomenal features, like color, tone, and smell qualia, remain outside of the explanatory reach of the model as it is; even though they all certainly often carry personal relevance as sensory qualities.

Another missing link in our theory is the status and role of language in the workspace, how it interacts with the mechanism of conscious access and how symbolic representations (e.g., words) and projective imagination or mental imagery mutually influence each other. This is an important question, but it is beyond the scope of this contribution. It will be the topic of a future communication, as we have specific hypotheses about this issue [1]. We note that active inference as a generalization of Bayesian inference can be set up to operate based on propositional models and logical reasoning.

Finally, the PCM does not tell us how subjectivity, sensory qualities, and representations with a phenomenal character relate, in their very phenomenal properties, to the computational and biophysical properties of the brain. Even though the model features a key relationship between computational efficacy and the mode of phenomenal access to spatial information in consciousness, and can be used to derive neurocomputational hypotheses, it neither solves the Hard Problem [32] nor closes the Explanatory Gap [147] completely. On the other hand, it could well be that the construction of mathematical models that allow one to intelligibly link the phenomenology of conscious experience to experimentally confirmed neurocomputational implementations of the models are all that one could or should reasonably expect in this regard. In the absence of compelling arguments for dualism or panpsychism, the neurophenomenological method suggested here can be taken as a key to the development of a physicalist theory of consciousness [see [4, 7, 8, 35, 51, 130, 148]. And even in the presence of such arguments, since only few would deny that every difference in consciousness corresponds to a difference in brain processing, the method and the resulting models retain their full significance.

## 4.5 Future directions for the model

Future versions of the kernel of our reduced-model will integrate a more differentiated and sophisticated repertory of distributions of personal relevance as related to shapes/morphologies, sets of basic actions, physical and social models. They will also integrate a full 3-dimensional projective geometry engine coupled to a virtual world containing meshes of environments, which will define a support for fields of conditional probabilities related to personal relevance. They will feature active inference of new structures in space to enrich the world-model through multiview reconstruction processing. The world-map itself can be dynamic and easily updated to create events that force the agent to a change in perspective. Complex structures can be modeled with attached personal relevance maps. The kernel will approach and avoid these structures accordingly: e.g., avoiding walls, approaching doors (an implicit collision avoidance mechanics is embedded in the kernel; it is all a matter of setting up probabilities).

Already with the reduced-model, the instantiated agents can be rendered more or less intelligent and resilient, depending on the tuning of several parameters such as the rate and duration of pure imagination-mode, the prior updating rate, and the initial conditions in the situation of the agents, in terms of both factual and believed given information, along with the agents' developmental history. But the capacity of their consciousness will be enhanced and enriched by explicitly adding in the model a layer of appraisal of self-relevance and self-concern based on a mechanism of perspective taking associated with the putative standpoint of other agents, i.e., a model of intersubjectivity based on the Theory of Mind. This will require to embed and instantiate virtual simulations of other agents' consciousnesses in the algorithm of the agent itself, based on the same principle used for the basic model.

We can envision using collectivities of such agents in interaction (they would be quite tractable computationally), which would have prior biases towards each other, e.g., some leading to mutual

attraction, others to mutual repulsion, and others to stalking behaviors. Agent 'liking' each other, could easily share their knowledge (prior maps) based on 'trust', and mutually update their prior maps through simple message-passing algorithms. In cases of disappointment, i.e., a mismatch between the information provided and actual sensory evidence, the probabilities of personal relevance attached to the agent source of information would naturally be negatively tuned down. Interesting social behaviors should emerge.

## 4.6   Conclusion and perspectives

The present Projective Consciousness Model contributes to the overarching goal of mathematizing phenomenology [35, 148]. One important incentive for deriving generative models of consciousness is the possibility of using such models as mathematical tools for the development of a phenomenological psychology, based on sound formal and computational foundations. If our model contains some truth, a future formal psychological science may include the study (and the classification) of possible conscious states, based on projective solutions from projective geometry in the context of processes of active inference driven by the minimization of free energy. Computational implementation of the PCM can be used to predict human behavior as a function of parameters across ranges of appraisal profiles, in normal and pathological contexts, with broad clinical perspectives. Another connected incentive for deriving generative models is to make connections with other fields of science that build generative models (physics, chemistry, biology, computer science), and to make connections with applications, including clinical models and applications. We believe that our model offers a promising toolkit and potential formal foundations for psychology, neuroscience and artificial intelligence.

# 5   Acknowledgments

# References

1. David Rudrauf. Structure-function relationships behind the phenomenon of cognitive resilience in neurology: insights for neuroscience and medicine. *Advances in Neuroscience*, 2014, 2014.

2. Edmund Husserl. *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy: First Book: General Introduction to a Pure Phenomenology*. Springer 1st edition, 1913.

3. Maurice Merleau-Ponty. *Phenomenology of perception*. New York: Routledge, 1945.

4. Kenneth Williford. *The self-representational structure of consciousness*. In Self-representational approaches to consciousness U. Kriegel and K. Williford (eds.). The MIT Press, 2006.

5. Kenneth Williford, David Rudrauf, and Gregory Landini. The paradoxes of subjectivity and the projective structure of consciousness. *Consciousness and Subjectivity*, 2012.

6. Francisco Varela. *Principles of Biological Autonomy*. New York: Elsevier/North-Holland, 1979.

7. Francisco J Varela. The specious present: A neurophenomenology of time consciousness. *Naturalizing phenomenology*, pages 266–314, 1999.

8. David Rudrauf, Antoine Lutz, Diego Cosmelli, Jean-Philippe Lachaux, and Michel Le Van Quyen. From autopoiesis to neurophenomenology: Francisco varela's exploration of the biophysics of being. *Biological research*, 36(1):27–65, 2003.

9. Antonio Damasio and David Rudrauf. *The Biological Basis of Subjectivity: A Hypothesis.* Self-Representational Approaches to Consciousness, Cambridge MA, MIT Press/Bradford Books, 2006.

10. Karl Friston. The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2):127–138, 2010.

11. Alain Berthoz. *Sens du mouvement (Le)*. Odile Jacob, 1997.

12. Marvin J Greenberg. *Euclidean and non-Euclidean geometries: Development and history.* Macmillan, 1993.

13. Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision.* Cambridge university press, 2003.

14. Olaf Blanke and Thomas Metzinger. Full-body illusions and minimal phenomenal selfhood. *Trends in cognitive sciences*, 13(1):7–13, 2009.

15. Bernard J Baars. Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. *Progress in brain research*, 150:45–53, 2005.

16. Bernard J Baars and Stan Franklin. Consciousness is computational: The lida model of global workspace theory. *International Journal of Machine Consciousness*, 1(01):23–32, 2009.

17. Giulio Tononi. An information integration theory of consciousness. *BMC neuroscience*, 5(1):42, 2004.

18. Stanislas Dehaene, Jean-Pierre Changeux, Lionel Naccache, Jérôme Sackur, and Claire Sergent. Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends in cognitive sciences*, 10(5):204–211, 2006.

19. Francis Crick and Christof Koch. A framework for consciousness. *Nature neuroscience*, 6(2):119–126, 2003.

20. Anil K Seth. The cybernetic bayesian brain. In *Open Mind*. Open MIND. Frankfurt am Main: MIND Group, 2014.

21. Anil K Seth. Presence, objecthood, and the phenomenology of predictive perception. *Cognitive neuroscience*, 6(2-3):111–117, 2015.

22. Jakob Hohwy. Elusive phenomenology, counterfactual awareness, and presence without mastery. *Cognitive neuroscience*, 5(2):127–128, 2014.

23. Andy Clark. Spreading the joy? why the machinery of consciousness is (probably) still in the head. *Mind*, 118(472):963–993, 2009.

24. Matthew AJ Apps and Manos Tsakiris. The free-energy self: a predictive coding account of self-recognition. *Neuroscience & Biobehavioral Reviews*, 41:85–97, 2014.

25. Jakub Limanowski and Felix Blankenburg. Minimal self-models and the free energy principle. 2013.

26. Sergio Paradiso and David Rudrauf. Struggle for life, struggle for love and recognition: the neglected self in social cognitive neuroscience. *Dialogues Clin Neurosci*, 14(1):65–75, 2012.

27. Carissa L Philippi, Justin S Feinstein, Sahib S Khalsa, Antonio Damasio, Daniel Tranel, Gregory Landini, Kenneth Williford, and David Rudrauf. Preserved self-awareness following extensive bilateral brain damage to the insula, anterior cingulate, and medial prefrontal cortices. *PloS one*, 7(8):e38413, 2012.

28. Antonio R Damasio. *The feeling of what happens: Body and emotion in the making of consciousness.* Houghton Mifflin Harcourt, 1999.

29. David Rudrauf, Olivier David, Jean-Philippe Lachaux, Christopher K Kovach, Jacques Martinerie, Bernard Renault, and Antonio Damasio. Rapid interactions between the ventral visual stream and emotion-related structures rely on a two-pathway architecture. *The Journal of Neuroscience*, 28(11):2793–2803, 2008.

30. Klaus R Scherer. The dynamic architecture of emotion: Evidence for the component process model. *Cognition and emotion*, 23(7):1307–1351, 2009.

31. Jeffrey A Gray. Brain systems that mediate both emotion and cognition. *Cognition & Emotion*, 4(3):269–288, 1990.

32. David J Chalmers. Facing up to the problem of consciousness. *Journal of consciousness studies*, 2(3):200–219, 1995.

33. Stanislas Dehaene, Michel Kerszberg, and Jean-Pierre Changeux. A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the National Academy of Sciences*, 95(24):14529–14534, 1998.

34. Walter J Freeman. Consciousness, intentionality and causality. *Journal of Consciousness Studies*, 6(11-12):143–172, 1999.

35. Masafumi Oizumi, Larissa Albantakis, and Giulio Tononi. From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0. *PLoS Comput Biol*, 10(5):e1003588, 2014.

36. Ned Block. How many concepts of consciousness? *Behavioral and brain sciences*, 18(02):272–287, 1995.

37. David Sander, Jordan Grafman, and Tiziana Zalla. The human amygdala: an evolved system for relevance detection. *Reviews in the Neurosciences*, 14(4):303–316, 2003.

38. Andrew J Elliot. The hierarchical model of approach-avoidance motivation. *Motivation and emotion*, 30(2):111–116, 2006.

39. Raffael Kalisch, Marianne B Müller, and Oliver Tüscher. A conceptual framework for the neurobiological study of resilience. *Behavioral and Brain Sciences*, 38:e92, 2015.

40. Thomas W Cronin. The visual ecology of predator-prey interactions. *Ecology of predator-prey interactions*, pages 105–138, 2005.

41. Andy Clark. Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(03):181–204, 2013.

42. David C Knill and Alexandre Pouget. The bayesian brain: the role of uncertainty in neural coding and computation. *Trends in neurosciences*, 27(12):712–719, 2004.

43. Alan Yuille and Daniel Kersten. Vision as bayesian inference: analysis by synthesis? *Trends in cognitive sciences*, 10(7):301–308, 2006.

44. Roger C Conant and W Ross Ashby. Every good regulator of a system must be a model of that system. *International journal of systems science*, 1(2):89–97, 1970.

45. Immanuel Kant. *Critique of pure reason.* Cambridge University Press, 1999.

46. Dorothée Legrand, Claudio Brozzoli, Yves Rossetti, and Alessandro Farne. Close to me: Multisensory space representations for action and pre-reflexive consciousness of oneself-in-the-world. *Consciousness and cognition*, 16(3):687–699, 2007.

47. Bigna Lenggenhager, Michael Mouthon, and Olaf Blanke. Spatial aspects of bodily self-consciousness. *Consciousness and cognition*, 18(1):110–117, 2009.

48. Jan Koenderink and Andrea van Doorn. The structure of visual spaces. *Journal of mathematical imaging and vision*, 31(2-3):171–187, 2008.

49. A Gurwitsch. *The field of consciousness.* Pittsburg, Pa: Duquesne University Press, 1964.

50. Charles S Peirce. *The Principles of Phenomenology.* In Philosophical Writings of Peirce, New York: Dover Publications, 1955.

51. Kenneth Williford. Representationalisms, subjective character, and self-acquaintance. In *Open MIND.* Open MIND. Frankfurt am Main: MIND Group, 2015.

52. G L McCormack. *Fusion and Binocularity.* In Borishs Clinical Refraction ed. Benjamin WJ, WB Saunders Company, Philadelphia, 2006.

53. Hofstadter Douglas. *I am a Strange Loop.* Basic Books, 2007.

54. Thomas Metzinger. *Being no one: The self-model theory of subjectivity.* mit Press, 2004.

55. Olaf Blanke. Multisensory brain mechanisms of bodily self-consciousness. *Nature Reviews Neuroscience*, 13(8):556–571, 2012.

56. Laura Machado do Nascimento. Noë, alva. out of our heads: Why you are not your brain, and other lessons from the biology of consciousness. *Principia*, 16(3):495, 2012.

57. Uriah Kriegel and Kenneth Williford. *Self-representational approaches to consciousness.* Mit Press, 2006.

58. Panayot Butchvarov. *Being qua being: a theory of identity, existence, and predication.* Indiana University Press, 1979.

59. Panayot K Butchvarov. Skepticism about the external world.

60. Dennett Daniel. *Consciousness explained.* Boston, MA: Little Brown, 1991.

61. David Hume. *A Treatise of Human Nature.* Oxford University Press, 1739.

62. Ludwig Wittgenstein. *Tractatus Logico-Philosophicus.* London: Routledge, 1922.

63. Jean-Paul Sartre. *The Transcendence of the Ego: A sketch for a phenomenological description.* Routledge, 1937.

64. Kenneth Williford. Pre-reflective self-consciousness and the autobiographical ego. in reading sartre: On phenomenology and existentialism.

65. Melita J Giummarra and G Lorimer Moseley. Phantom limb pain and bodily awareness: current concepts and future directions. *Current Opinion in Anesthesiology*, 24(5):524–531, 2011.

66. Jonathan Smallwood and Jonathan W Schooler. The restless mind. *Psychological bulletin*, 132(6):946, 2006.

67. Hermann Schone. *Spatial orientation: the spatial control of behavior in animals and man*. Princeton University Press, 2014.

68. Claus Lamm, C Daniel Batson, and Jean Decety. The neural substrate of human empathy: effects of perspective-taking and cognitive appraisal. *Journal of cognitive neuroscience*, 19(1):42–58, 2007.

69. Nicholas Humphrey. *The inner eye: Social intelligence in evolution*. Oxford University Press, 2003.

70. Laura A Kelley and John A Endler. Illusions promote mating success in great bowerbirds. *Science*, 335(6066):335–338, 2012.

71. Jan Koenderink, Andrea van Doorn, Huib de Ridder, and Stijn Oomes. Visual rays are parallel. *Perception*, 39(9):1163–1171, 2010.

72. Henri Poincaré. *La science et l'hypothèse*. Booklassic, 2015.

73. William Wilkerson. Time and ambiguity: Reassessing merleau-ponty on sartrean freedom. *Journal of the History of Philosophy*, 48(2):207–234, 2010.

74. Andre M Bastos, W Martin Usrey, Rick A Adams, George R Mangun, Pascal Fries, and Karl J Friston. Canonical microcircuits for predictive coding. *Neuron*, 76(4):695–711, 2012.

75. Rajesh PN Rao and Dana H Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, 2(1):79–87, 1999.

76. Karl Friston, Thomas FitzGerald, Francesco Rigoli, Philipp Schwartenbeck, and Giovanni Pezzulo. Active inference: A process theory. *Neural Computation*, 2016.

77. Karl Friston, Rick Adams, Laurent Perrinet, and Michael Breakspear. Perceptions as hypotheses: saccades as experiments. *Frontiers in psychology*, 3:151, 2012.

78. Karl Friston, Francesco Rigoli, Dimitri Ognibene, Christoph Mathys, Thomas Fitzgerald, and Giovanni Pezzulo. Active inference and epistemic value. *Cognitive neuroscience*, 6(4):187–214, 2015.

79. M Berk Mirza, Rick A Adams, Christoph D Mathys, and Karl J Friston. Scene construction, visual foraging, and active inference. *Frontiers in Computational Neuroscience*, 10, 2016.

80. Laurent Itti and Pierre Baldi. Bayesian surprise attracts human attention. *Vision research*, 49(10):1295–1306, 2009.

81. Karl Friston. Hierarchical models in the brain. *PLoS Comput Biol*, 4(11):e1000211, 2008.

82. John O Campbell. Universal darwinism as a process of bayesian inference. *Frontiers in Systems Neuroscience*, 10, 2016.

83. Rick A Adams, Stewart Shipp, and Karl J Friston. Predictions not commands: active inference in the motor system. *Brain Structure and Function*, 218(3):611–643, 2013.

84. Anil K Seth and Karl J Friston. Active interoceptive inference and the emotional brain. *Phil. Trans. R. Soc. B*, 371(1708):20160007, 2016.

85. Lisa Feldman Barrett and W Kyle Simmons. Interoceptive predictions in the brain. *Nature Reviews Neuroscience*, 16(7):419–429, 2015.

86. Harriet Feldman and Karl Friston. Attention, uncertainty, and free-energy. *Frontiers in human neuroscience*, 4:215, 2010.

87. Emil Artin. *Geometric algebra*. Courier Dover Publications, 2016.

88. Andy Clark. *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press, 2015.

89. Jakob Hohwy. *The predictive mind*. Oxford University Press, 2013.

90. Gerrans P Brosch T Sander D Murray, R J. When at rest: event-free active inference may give rise to implicit self-models of coping potential. *Behavioral and Brain Sciences*, 38(e114):41–42, 2015.

91. Karl Friston. A theory of cortical responses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 360(1456):815–836, 2005.

92. William A Cunningham, Kristen A Dunfield, and Paul E Stillman. Emotional states from affective dynamics. *Emotion Review*, 5(4):344–355, 2013.

93. Mateus Joffily and Giorgio Coricelli. Emotional valence and the free-energy principle. *PLoS Comput Biol*, 9(6):e1003094, 2013.

94. Daniel Bennequin. Remarks on invariance in the primary visual systems of mammals. In *Neuromathematics of Vision*, pages 243–333. Springer, 2014.

95. Jean-Paul Sartre. *The imagination*. London: Routledge.(Translation Williford, K., Rudrauf, D.), 2012.

96. Jean-Paul Sartre. *The imaginary: A phenomenological psychology of the imagination*. Routledge Press, 2004.

97. Colin McGinn and Colin McGinn. *Mindsight: Image, dream, meaning*. Harvard University Press, 2009.

98. Amy Kind. *The Routledge Handbook of Philosophy of Imagination*. Routledge, 2016.

99. Kenneth Williford. Husserls hyletic data and phenomenal consciousness. *Phenomenology and the cognitive sciences*, 12(3):501–519, 2013.

100. Jim Hopkins. Free energy and virtual reality in neuroscience and psychoanalysis: A complexity theory of dreaming and mental disorder. *Frontiers in Psychology*, 7, 2016.

101. Jacques Droulez and Alain Berthoz. The concept of dynamic memory in sensorimotor control. *Motor control: Concepts and issues*, pages 137–161, 1991.

102. Wilder Penfield and Edwin Boldrey. Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain: A journal of neurology*, 1937.

103. Agnes Moors and Klaus R Scherer. The role of appraisal in emotion. *Handbook of cognition and emotion*, pages 135–155, 2013.

104. William James. *The principles of psychology*. Read Books Ltd, 1913.

105. Morris Rosenberg. Society and the adolescent self-image. 1965.

106. Jacques Lacan and Bruce Fink. *Ecrits: A selection*. WW Norton & Company, 1966,2002.

107. John F Kihlstrom and Nancy Cantor. Mental representations of the self. *Advances in experimental social psychology*, 17:1–47, 1984.

108. A Kappas. The sociality of appraisals: Impact of social situations on the evaluation of emotion antecedent events and physiological and expressive reactions. In *1SRE 1996: Proceedings of the IXth conference of the International Society for Research on Emotions*, pages 116–120, 1996.

109. Antony SR Manstead and Agneta H Fischer. Social appraisal: The social world as object of and influence on appraisal processes. *Appraisal processes in emotion: Theory, methods, research*, pages 221–232, 2001.

110. Harold Scott Macdonald Coxeter. Introduction to geometry. 1961.

111. Chai-Youn Kim and Randolph Blake. Psychophysical magic: rendering the visible invisible. *Trends in cognitive sciences*, 9(8):381–388, 2005.

112. Marc K Albert and Donald D Hoffman. The generic-viewpoint assumption and illusory contours. *Perception*, 29(3):303–312, 2000.

113. GM Stratton. The spatial harmony of touch and sight. *Mind*, 8(4):492–505, 1899.

114. H Henrik Ehrsson. The experimental induction of out-of-body experiences. *Science*, 317(5841):1048–1048, 2007.

115. Halim Hicheur, Hideki Kadone, Julie Grezes, and Alain Berthoz. Perception of emotional gaits using avatar animation of real and artificially synthesized gaits. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, pages 460–466. IEEE, 2013.

116. Simon Lambrey, Christian Doeller, Alain Berthoz, and Neil Burgess. Imagining being somewhere else: neural basis of changing perspective in space. *Cerebral cortex*, 22(1):166–174, 2012.

117. Bérangère Thirioux, Manuel R Mercier, Gérard Jorland, Alain Berthoz, and Olaf Blanke. Mental imagery of self-location during spontaneous and active self–other interactions: An electrical neuroimaging study. *The Journal of Neuroscience*, 30(21):7202–7214, 2010.

118. Dan Zahavi. *Subjectivity and selfhood: Investigating the first-person perspective.* MIT press, 2008.

119. D Henrich. Fichtes original insight. *Contemporary German Philosophy*, 1:15–53, 1966.

120. Dan Zahavi. *Self-awareness and alterity: A phenomenological investigation.* Northwestern University Press, 1999.

121. Manfred Frank. Fragments of a history of the theory of self-consciousness from kant to kierkegaard. *Critical Horizons*, 5(1):53–136, 2004.

122. Eric Klinger. *Structure and functions of fantasy.* Wiley-Interscience, 1971.

123. Reineberg Andrew E Schooler Jonathan W Reichle, Erik D. Eye movements during mindless reading. *Psychological Science*, 21(9):1300–1310, 2010.

124. Heinrich Klüver. Mescal visions and eidetic vision. *The American journal of psychology*, 37(4):502–515, 1926.

125. Heinrich Klüver. *Mescal, and mechanisms of hallucinations.* [Chicago]: University of Chicago Press, 1966.

126. Christopher W Tyler. Some new entoptic phenomena. *Vision research*, 18(12):1633–1639, 1978.

127. Yves Fregnac. Neurogeometry and entoptic visions of the functional architecture of the brain. *Journal of Physiology-Paris*, 97(2):87–92, 2003.

128. Ronald K Siegel. *Intoxication: The universal drive for mind-altering substances.* Inner Traditions/Bear & Co, 2005.

129. Thomas Metzinger. Out-of-body experiences as the origin of the concept of a'soul'. *Mind and Matter*, 3(1):57–84, 2005.

130. Francisco J Varela. Neurophenomenology: A methodological remedy for the hard problem. *Journal of consciousness studies*, 3(4):330–349, 1996.

131. Arturo Tozzi and James F Peters. Towards a fourth spatial dimension of brain activity. *Cognitive neurodynamics*, 10(3):189–199, 2016.

132. Salma Mesmoudi, Vincent Perlbarg, David Rudrauf, Arnaud Messe, Basile Pinsard, Dominique Hasboun, Claudia Cioli, Guillaume Marrelec, Roberto Toro, Habib Benali, et al. Resting state networks' corticotopy: the dual intertwined rings architecture. *PloS one*, 8(7):e67444, 2013.

133. John O'Keefe, Neil Burgess, James G Donnett, Kathryn J Jeffery, and Eleanor A Maguire. Place cells, navigational accuracy, and the human hippocampus. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 353(1373):1333–1340, 1998.

134. Peter Zeidman, Antoine Lutti, and Eleanor A Maguire. Investigating the functions of subregions within anterior hippocampus. *cortex*, 73:240–256, 2015.

135. Takuma Kitanishi, Hiroshi T Ito, Yuichiro Hayashi, Yoshiaki Shinohara, Kenji Mizuseki, and Takatoshi Hikida. Network mechanisms of hippocampal laterality, place coding, and goal-directed navigation. *The Journal of Physiological Sciences*, pages 1–12, 2016.

136. Randy L Buckner, Jessica R Andrews-Hanna, and Daniel L Schacter. The brain's default network. *Annals of the New York Academy of Sciences*, 1124(1):1–38, 2008.

137. Leslie G Ungerleider and James V Haxby. whatand wherein the human brain. *Current opinion in neurobiology*, 4(2):157–165, 1994.

138. Leslie G Ungerleider. Two cortical visual systems. *Analysis of visual behavior*, pages 549–586, 1982.

139. Paolo Bartolomeo and Sylvie Chokron. Orienting of attention in left unilateral neglect. *Neuroscience & Biobehavioral Reviews*, 26(2):217–234, 2002.

140. Richard A Andersen. Multimodal integration for the representation of space in the posterior parietal cortex. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 352(1360):1421–1428, 1997.

141. Jeffrey M Zacks. Neuroimaging studies of mental rotation: a meta-analysis and review. *Journal of cognitive neuroscience*, 20(1):1–19, 2008.

142. Demis Hassabis and Eleanor A Maguire. Deconstructing episodic memory with construction. *Trends in cognitive sciences*, 11(7):299–306, 2007.

143. Simone Vossel, Christoph Mathys, Klaas E Stephan, and Karl J Friston. Cortical coupling reflects bayesian belief updating in the deployment of spatial attention. *Journal of Neuroscience*, 35(33):11532–11542, 2015.

144. Jason Yosinski, Jeff Clune, Anh Nguyen, Thomas Fuchs, and Hod Lipson. Understanding neural networks through deep visualization. *arXiv preprint arXiv:1506.06579*, 2015.

145. Bruno Herbelin, Frederic Vexo, and Daniel Thalmann. Sense of presence in virtual reality exposures therapy. In *Proceedings of the 1st International Workshop on Virtual Reality Rehabilitation, Lausanne, Switzerland*. Citeseer, 2002.

146. Sahib S Khalsa, David Rudrauf, Justin S Feinstein, and Daniel Tranel. The pathways of interoceptive awareness. *Nature neuroscience*, 12(12):1494–1496, 2009.

147. Joseph Levine. Materialism and qualia: The explanatory gap. *Pacific philosophical quarterly*, 64(4):354–361, 1983.

148. Jean Petitot, Francisco J Varela, Bernard Pachoud, and Jean-Michel Roy. *(Eds.) Naturalizing phenomenology: Issues in contemporary phenomenology and cognitive science*. Stanford University Press, 1999.