



# Explaining “spatial purport of perception”: a predictive processing approach

Wiktor Rorot<sup>1</sup> 

Received: 8 November 2019 / Accepted: 27 April 2020  
© The Author(s) 2020

## Abstract

Despite the large interest in the human ability to perceive space present in neuroscience, cognitive science and psychology, as well as philosophy of mind, the issues regarding egocentric space representation received relatively less attention. In this paper I take up a unique phenomenon related to this faculty: the “spatial purport” of perceptual experiences. The notion was proposed by Rick Grush to describe the subjective, qualitative aspects of egocentric representations of spatial properties and relations. Although Grush offered an explanation of the mechanism giving rise to appearance of spatial purport, his model had considerable shortcomings. In the paper I thoroughly analyze both the notion of spatial purport and Grush’s explanation of the mechanism at its core in order to develop his theory using the insights provided by the predictive processing theory of mind, and more particularly by the active inference framework. The extended account I offer, named Predictive and Hierarchical Skill Theory, explains phenomena that escaped Grush’s model and furthers the research on egocentric space representation from the perspective of both neuroscience and philosophy of mind.

**Keywords** Egocentric space representation · Spatial purport · Predictive processing · Active inference · Free energy principle · Space perception · Egocentric space

## 1 Introduction

Historically speaking, issues related to the human ability to perceive space have received uneven treatment from philosophers, cognitive scientists and psychologists alike. Since the discovery that this ability is driven by two distinct and immensely complex systems, one responsible for the so-called allocentric and the other for the egocentric space representations, scientists have focused on the former. This research has led to a clear understanding of the hippocampal systems of spatial representation

---

✉ Wiktor Rorot  
wiktor.rorot@student.uw.edu.pl

<sup>1</sup> Institute of Philosophy, University of Warsaw, Warsaw, Poland

(see Moser et al. 2008). However, the other system still remains mysterious. A plausible explanation of this state of affairs seems to come from the close relation between egocentric spatial representation and elusive subjective and qualitative aspects of the human experience of space. To be more precise: the egocentric representation of space is tied to everyday phenomenal experiences such as the feeling of being located somewhere, being distant from other physical objects, being oriented towards one particular side.

Rick Grush—and Gareth Evans before him—have attempted to address those specific phenomena from an embodied and embedded perspective (Evans 1985; Grush 1998, 2000, 2007b, 2009). Grush offered an interesting account of the mechanism underlying them, the “Skill Theory v2.0” (Grush 2007b). Unfortunately, the original formulation of his proposition explicitly excludes an important aspect of space perception, object-motions, despite the fact that the emulation framework Grush was concurrently developing (Grush 2004) did offer necessary tools to explain it. What’s more, although the Skill Theory is currently a state-of-the-art model, it did not prompt empirical studies.

In this paper I will try to further develop this model. The goal is to offer a comprehensive description of the neural mechanism underlying egocentric space perception, reformulating Grush’s original account. Hopefully, this formulation will operationalize the studied phenomenon in such a way as to enable future experimental examination. To this purpose I will turn to the predictive processing framework (Clark 2016; Wiese and Metzinger 2017, henceforth PP), that will serve as the bedrock for the proposed model, named Predictive and Hierarchical Skill Theory—PHiST.

In the next sections I will discuss how Grush describes the phenomenology of egocentric space perception with the notion of “spatial purport of perception”, how the term can be defined, and how the skill theory v2.0 describes the neural mechanism underlying this phenomenon. Then, I will show why Grush’s model does not offer a sufficient description and how it can be rebuilt into PHiST, following insights from the predictive processing theory. Finally, I will summarize this discussion, showing how PHiST can help us better understand what spatial purport actually is.

## 2 Spatial purport of perception

Gareth Evans, discussing the relations between spatial concepts<sup>1</sup> of different sensory modalities, advocates to regard them as identical across the senses. He claims that this identity is rooted in behavioral dispositions underlying space perception and illustrates this point with an analysis of how humans discern the directionality of sounds:

---

<sup>1</sup> The term “spatial concept” appears in discussions of perceptual experience mainly for historical reasons and has a specific, non-standard meaning, which underscores the significance of qualitative aspects associated with these concepts: the experiences of breadth, width, and other spatial properties of physical objects in environment. In this paper I will mainly, and at times anachronically, use the notion of spatial *purport*, however, at some points it will be necessary to refer to spatial *concepts*. At these points the reader should bear in mind that what is meant involves this phenomenal aspect and so these concepts involve purport. I am thankful to an anonymous reviewer for prompting me to clarify this terminological issue.

We do not hear a sound as coming from a certain direction, and then have to *think* or calculate which way to turn our heads to look for the source of the sound etc. [...] [H]ow is that position to be specified? We envisage specifications like this: [the subject] hears the sound up, or down, to the right [...] It is clear that these terms are *egocentric* terms [that] derive their meaning from their (complicated) connections with the actions of the subject. (Evans 1985, pp. 383–384)

Evans termed this view “disposition theory” since the spatial features of our perceptual experience arise from the dispositions to act they prompt and as such are inherent parts of our phenomenal experience. For Evans spatial concepts also arise, necessarily, from such behavioral dispositions, building upon these qualities.<sup>2</sup>

Rick Grush has taken up the issues related to space perception from the perspective offered by Evans. The skill theory is a rendering of the disposition theory in terms of contemporary cognitive science: in the form of a computational model sensitive to a possible neural implementation. The theory is centered on this phenomenology that was as well Evans’ focus. To account for it Grush coins the term *spatial purport of perception* (Grush 2007b, p. 390). Before I will turn to a presentation of the skill theory v2.0, it is necessary to review what Grush means by this notion: what is the explanandum here.

## 2.1 Purport, content, phenomenology...

Grush follows Evans and returns to the early modern philosophers, most significantly George Berkeley, and thoroughly analyzes their views on (visual) perception (Grush 2007a). Berkeley did not accept that any spatial concept could be shared between sensory modalities. He offered an argument to show that spatial features of visual perception are not “proper objects of vision”, but are derived from quasi-spatial features together with genuinely spatial experiences of touch, kinesthesia, and proprioception.

This argument highlights the difference between “carrying information about space” (note that information is meant here purely in the technical sense of Shannon information) and “having spatial purport” (Grush 2007a, pp. 427–428). What is the case is that vision, for Berkeley, via the numerous quasi-spatial manifolds, carries information about space—i.e. the amount of “eye strain” necessary to sharply see an object is the information about the distance of the object from the subject. But in itself, it has no spatial purport: “the manifolds are not intrinsically coordinated [...] at least before relevant experience” and “any learned coordinations [between quasi-spatial manifolds of vision and genuinely spatial manifolds of touch] are contingent” (Grush 2007a, p. 428). However, Berkeley’s account is obviously uninformative with regard to what the purport is.

<sup>2</sup> The disposition theory bears significant similarity to the theory of sensorimotor contingencies, as developed by e.g., (Barsalou 1999; O’Regan and Noë 2001), which on the other hand has been integrated with the predictive processing framework (Seth 2014). There has been also some work on the relation between sensorimotor contingencies and the perception of space, however, it focused on allocentric space representation (Konig et al. 2016). Grush however explicitly discards sensorimotor contingencies as a way to account for spatial purport, since they do not provide an account of representation of the state of the world (Grush 2007b, p. 413). Examination of the relations between these accounts is outside of the scope of current paper.

To my knowledge, the most extensive presentation of this concept in Grush's work is the following:

My concern is the spatial content of perceptual experience. But since the word "content" often is used to indicate what some word or mental state is about, it won't quite suit my purposes. As I will explain later, there can be states, in particular experiential states, that carry information about space [...], while not having any spatial significance for the subject. This makes it sound like I am interested in spatial phenomenology, and I think this is right. But again, the word "phenomenology" and its various cognates are very loaded. I don't want to get mired in an argument as to whether there are "spatial" qualia, for example. I will use the expression "purport" [...] to indicate what it is I am after. I will give a fuller characterization of what I mean by "purport" shortly. If it turns out that what I mean by purport is what you mean by phenomenology, or content, then fine by me. (Grush 2007b, p. 390)

But the promise remains unfulfilled, as the "fuller characterization" later in the text amounts to an example that can be treated at most as a description of the core phenomenon his theory aims to explain. The example is following: Consider two people using a "sonic guide", a device that translates spatial properties of environment into auditory cues. One of them, let us call her Inga, is congenitally blind and has been using the guide for many years. The other person, Otto, is sighted and has been blinded and introduced to the guide only for the purpose of the experiment. What will those two people perceive, when they will be introduced into a new environment (e.g. a cluttered room)? While Inga will be able to navigate using the sounds she hears, Otto will be completely lost, unable to move, presented by the guide with a meaningless cacophony of sounds. Grush claims that exactly this difference is what he attempts to capture by the notion of "purport": for Inga, the guide will give rise to genuinely spatial experience, while for Otto it will not, despite the fact that in both cases it will carry the same (or fairly similar) information about space (again, in the technical sense of Shannon information).

Further in his article Grush complicates this example: imagine now that Otto is a musical genius, gifted with the absolute pitch, who swiftly learns the complicated dynamics of a sound getting higher and louder as he approaches an obstacle. But even then, when he will learn the sounds' internal dynamics, Grush claims, nothing will change in his experience—sounds have for him no "spatial significance", while for Inga they do. If Otto is to have genuine spatial purport via the device, he truly has to master it (learning the relation between the sounds and the state of the world, as described by the disposition theory).

This is far from a sufficient presentation of what is meant by spatial purport. However, this description hints at some of the elements that spatial purport comprises of. I will try to examine them, before turning to an attempt to offer a succinct definition of the concept.

## 2.2 Elements of spatial purport

Spatiality of the perceived world amounts to objects being at some distance from each other, including the distance from the subject, and the distance between their parts. It also includes directionality—the fact that turning your head to the left and to the right are two distinct moves, usually enabling different objects to enter your visual field. An important aspect of spatiality of experience consists of the phenomenal “point of view”, that is usually experienced as located within one’s body, and functions as a perspective for all perceptual experiences (most notably for visual perception). What’s more, this point of view can be perceived as located also within some more general space, that is—being spatially related to the objects of perception.

This is due to the human ability to formulate self-location judgments based on their sensory experience, even if the exact relation between the contents of experience and the judgments is in this case subject to debate (see Schwenkler 2014). Although this debate focuses on visual experience and disregards all other sensory modalities, we can make use of the conceptual distinctions it introduces to throw more light on spatial purport.

### 2.2.1 Self-locating contents of perception

The disagreement on the epistemological role of experience for the self-location judgments can be seen as a repetition of Berkeley’s argument outlined in the previous section. Proponents of the Minimal View, i.e. the view that a person’s visual experience doesn’t involve, in virtue of its perspectival nature, any representation of where they are located with respect to their surroundings (Schwenkler 2014, p. 141) would most likely agree with Berkeley’s claim that spatial properties are not “proper objects of vision”, and hence, vision at most can carry information about space. On the other hand, proponents of the Self-Location Thesis claim that “simply in virtue of its perspectival character, visual experience can include the location of the perceiver among its face value contents” (Schwenkler 2014, p. 139). If we were to put their claim in the language that Grush adheres to, we would say that they claim that there is genuine spatiality in visual perception, that it has spatial purport.

Once we expand our analysis to all sensory modalities that undoubtedly do carry spatial information (disregarding their phenomenological contents, for now), namely: touch, interoception and kinetoception, vision, and hearing, the Self-Location Thesis becomes much easier and more straightforward to defend. A multisensory experience of the world does, in its totality, include the experience of being located within the world experienced. I do not have to think of whether I am sitting in front of the desk, facing the tree outside of the window. I experience myself as being so located within the world. At the same time, this remains a purely egocentric experience, as even when I think of objects occluded by others from my current perspective, objects I do not directly perceive, I think of them in those egocentric terms, as located to my left or to my right. Here, I believe, John Schwenkler is incorrect in characterizing the egocentric experience as “centered on a self, ignorant of itself” (Schwenkler 2014, p. 153),<sup>3</sup> since

<sup>3</sup> Schwenkler assigns this view of egocentric experience to Jean Piaget, however his characterization of Piaget’s views is overly simplified.

the self is included in the perspectival character of sensory experience, especially once we consider interoceptive experiences.

Self-location seems to be the foundation of what Grush calls the spatial purport. If we were to perceive only spatial relations between external objects without perceiving these objects as somehow spatially related to ourselves, e.g. in some elaborate afterimage experience where we perceive an array of precisely defined shapes but are unable to pin down their location in the world, our experience would not have neither spatial purport, nor self-locating contents. It seems then to be a necessary condition for the spatial purport to arise.

However, for our experience to have self-locating contents, there are two conditions that have to be met. First of all, there has to be a self that is being located by this experience. This is provided in the perspectival character of our experience: the “self” is just the starting point of our (egocentric) frame of reference (Evans 1982, pp. 153–154). (The self is conceived here in this minimal sense of a point of view towards the world, without presupposing its substantiality. E.g. Thomas Metzinger’s “phenomenal self” and the notion of “transparent self-model” (Metzinger 2004) would be sufficient for current purposes.) Second of all, there have to be some objects of perception that are experienced as standing in some relation to the perceiver. This relation is necessary for self-location. But to perceive objects as standing in some relation to the subject, as Quassim Cassam argues, the subject has to be “intuitively aware of oneself [...] as shaped, located and solid” (Cassam 2005, p. 52). There is, then, a circularity in the experience of self-location, circularity that hinges on the relation between the subject and the objects of perception.

However, once we think of the subject as a living, behaving organism, this relation can be thought of differently, without any immediate reference to spatial terms. Namely, we can define this relation as the ability to act in some manner on the objects of perception, referring back to Evans’ dispositions. In result, this circularity in the self-locating contents of perception boils down to the dependence on our bodily dispositions—or skills—to act on the environment we perceive.

### 2.3 Towards a definition

As I have argued above, spatial purport certainly involves experiences of *being distant from objects*, of *directionality*, of *self-location*. But the experience of being distant (or of distance between objects) and of directionality are circularly co-dependent on self-location, as there has to be a spatial self to perceive objects as distant or directed, and there have to be external, spatially located objects for there to be such a self. Hence, we can attempt to define those experiences only in their totality which is grounded in our bodily skills, or dispositions to act.

Spatial purport of perceptual experience is, hence, constituted by those phenomenal and qualitative aspects of the content of perceptual experience which we normally understand as experiences of distance, direction, self-location. They are inherently dependent on our embodiment and embeddedness within the physical world, as they are necessarily tied to our dispositions to act. But there is, again, a co-dependence, when we move from the level of skills, or action-types, to the level of particularly

enacted action-tokens, as our skillful performance depends precisely on the spatial contents of our experience.

This final co-dependence is what motivates, for the purposes of this project, the shifts between the phenomenological language describing phenomenal experience of space and the vocabulary of space representation. These shifts are based on a physicalist assumption: if one accepts a non-epiphenomenal view of consciousness, independently of whether it is treated as real or illusory, it has to be associated with some functional role in the cognitive system and hence is most likely realized by the same neural mechanism as the one that realizes this functional role. As a result, we may operationalize thinking of spatial purport of perception in terms of egocentric representation of space, as these representations are involved in guiding our particular action-tokens.<sup>4</sup>

### 3 Skill theory v2.0

Grush's skill theory offers a substantive revision and update of Evans' disposition theory. Grush at one point explains his account in the form of an equation: "Disposition theory plus trajectory emulation theory = Skill Theory v2.0" (Grush 2007b, p. 405). More precisely, the Evansian disposition theory is here rendered plausible from the perspective of neuroscience, via a reference to the *basis function model*, a computational model of the posterior parietal cortex, a neural region largely responsible for egocentric spatial representation (Buneo and Andersen 2006; Zipser and Andersen 1988; Pouget and Sejnowski 1997; Pouget et al. 2002). The trajectory emulation theory is, on the other hand, a specific implementation of the more general emulation framework previously offered by Grush (2004), that specifies the formal methods employed for the description of the mechanism of spatial purport.

#### 3.1 Disposition theory and basis functions

A full depiction of the basis function model is here unnecessary [the interested reader is encouraged to reach to the original papers discussing it (Pouget and Sejnowski 1997; Pouget et al. 2002)]. Nevertheless, some basic features need to be introduced. First of all, the model aims to describe the operation of the posterior parietal cortex (henceforth PPC) which is regarded in the literature as the crucial cortical area responsible for the egocentric representation of space in humans and primates (Buneo and Andersen 2006). In some cases PPC lesions may also cause an inability to integrate the somatosensory information about the body's position with the visual inputs (Zipser and Andersen 1988).

The PPC has inputs from the visual, auditory and somatosensory systems, as well as from the motor and premotor cortices (Gharbawie et al. 2011). There is a consensus about its role as a sensorimotor interface (Andersen and Buneo 2003; Buneo and

<sup>4</sup> Note that in Grush's work occurs an analogous, although opposite and implicit shift: namely, in his (2004) paper when he refers to space perception, he does so using representational language, while in (2007b) presentation of the Skill Theory the representational vocabulary gives way to the phenomenological one, without a specification of the relation between the two.



Andersen 2006), as well as about the fact that while it is used to represent space, it differs from the hippocampus in that there are no topographic representations (such as hippocampal place cells, cf. O’Keefe 1998; Hafting 2005; Moser et al. 2008).

The basis function model (Zipser and Andersen 1988; Pouget and Sejnowski 1997; Pouget et al. 2002) was proposed as a computational schema of the recoding and remapping between data from different sensory modalities incoming to the PPC. Such a recoding is necessary since each modality represents the stimuli in its own frame of reference, e.g. the visual system uses the eye-centered map (due to the topography of retina), while the auditory system “prefers” the head-centered frame that arises from the processing performed early in the auditory cortex (Pouget and Sejnowski 1997; Pouget et al. 2002). What’s more, all cognitive and motor functions have preferred frames of reference of their own that they impose on the sensory data. Hence, to explain how it is possible that data from two or more modalities are used jointly to direct motion, it must be understood how are they interfaced between different frames of reference.

David Zipser and Richard Andersen (1988) attempted to model the coordinate transformation performed by PPC with a back-propagation trained artificial neural network, concluding that it is actually possible for the PPC to switch between different coordinate systems. Their work was continued by Alexandre Pouget, who proposed a basis function model that accounts for the operations performed by the hidden layer of an ANN similar to the one proposed by Zipser and Andersen. Pouget’s network (Pouget et al. 2002) has two input layers (one for eye-centered, i.e. retinal, location stimuli, and one for eye position), one hidden layer and an output layer that represents the location of stimuli in head-centered coordinates. The “head-centered units” value (marked as **O**) is not a simple addition of retinal location (**R**) and eye position (**E**), but a nonlinear (Gaussian, sigmoid or Gaussian-sigmoid) combination of **R** and **E**. Such nonlinear transformations can be approximated as linear combinations of basis functions.<sup>5</sup>

Pouget proposes that parietal neurons actually behave like  $b_i(\mathbf{S}, \mathbf{P})$ : the basis function of sensory inputs (**S**, a generalization of the retinal location **R**) and postural inputs (**P**, a generalization of the eye position **E**). If so, than the general operation of PPC may be described by the following equation (Pouget and Sejnowski 1997):

$$M^g = \sum_i c_i^g b_i(\mathbf{S}, \mathbf{P}) \quad (1)$$

Here,  $M^g$  is the motor command that PPC attempts to coordinate (with  $g$  indicating type of the movement, such as e.g. a grasp) and  $c_i^g$  are coefficients proprietary of a movement type  $g$  being computed.

<sup>5</sup> In linear algebra, when referred to a function space, a basis is a (finite or infinite) set  $B = \{b_i\}_{i \in I}$  of functions  $b_i$  (indexed by some appropriate index set  $I$ ) that are linearly independent and span the whole space—that is each function in that space can be expressed as a linear combination of basis elements of the form:

$$f = a_1 b_{i_1} + a_2 b_{i_2} + \dots + a_n b_{i_n}$$

where  $a_k$  are scalars—coordinates of the function  $f$  with respect to the basis  $B$ .



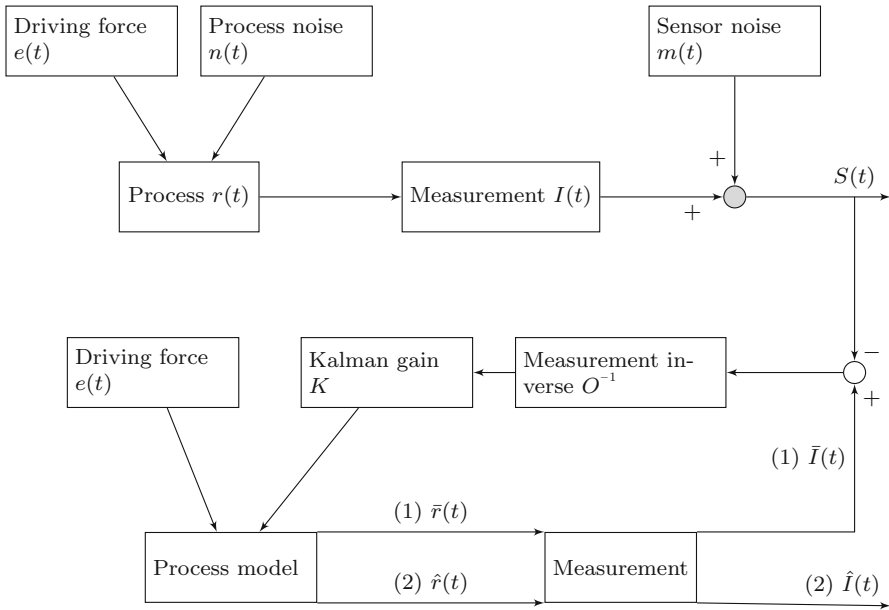


Fig. 1 A simple Kalman filter. See Grush (2004, pp. 380–381) for a detailed description

This is an extremely important observation, which supports Evans disposition theory: the operation of the system responsible for the egocentric space representation are on their basic level dependent on the agent’s dispositions to act, namely the motor commands  $M^s$ . Hence, Grush treats the basis function model as a neural implementation of the more abstract Evansian disposition theory. Basis functions are representations of the subject and objects they perceive within the world. But as Grush points out, this system describes only the static percepts and it needs to be expanded to cover the perception of spatial features extended in time.

### 3.2 Emulation framework

Grush’s “emulation theory of representation” (2004) offers an appropriate addition. It attempts to account for the mental faculties, such as motor control and imagery, perception, reasoning, theory of mind and language (Grush 2004) with a scheme based on control theory notions such as forward models and Kalman filters.<sup>6</sup> A Kalman filter (henceforth KF) is an algorithm for estimating values of unknown variables from a series of measurements containing noise and other inaccuracies (Zarchan and Musoff 2000). A standard (perhaps slightly simplified) KF is shown in the Fig. 1.

For Grush, a large amount of cognitive abilities, including the emergence of spatial purport of perception, can be specified in the following way:

<sup>6</sup> In this paper I will assume that the reader is acquainted with some of the basic cybernetic notions. Grush (2004) offers a friendly introduction to those ideas for those who seek a deeper understanding.

The general structure of the problem [is] quite unmysterious. One system (a ship's crew, a brain) is interacting with another system (a ship, an environment, a body<sup>7</sup>) such that the general principles of how this system functions are known, but the system's state is not entirely predictable owing to process noise (unpredictable currents, bodily or environmental perturbations), and imperfect sensors. The solution is to maintain a model of the process—done by a part of the ship's crew, the navigation team; or specialized emulator circuits in the brain—in order to provide predictions about what its state will be; and to use this prediction in combination with sensor information in order to maintain a good estimate of the actual state of the system. (Grush 2004, p. 382)

In case of spatial purport, the Kalman filter emulates future values of the basis functions of the sensory input.<sup>8</sup> It does so, since it knows the *driving force*—the future motor command  $M^s$  and maintains a *process model*, a forward model of subjects' and their surroundings' behavior in response to the action taken as a result of the command. Hence, if the PPC is equipped in the amodal emulator, it is able to predict future values of basis functions, namely future states of the world, rendering its operation much more dynamic. This explains the appearance of spatial purport in a changing environment, where this dynamics is a result of subjects' own actions.

#### 4 Skill theory v3.0, or PHiST

Unfortunately, Grush's skill theory v2.0, as it is presented in the (2007b) paper, is incomplete. The important shortcoming, one that the author is perfectly aware of, concerns perception of object-movement, movement generated by causes external to the subject. He explicitly states, that

[the] anticipations of object motion through behavioural space, produced as they are employing only [amodal emulator of basis functions]  $V_n$ , are limited to providing estimates of object trajectories that result from self-movement, and hence are predictions about movement in behavioral space. The trajectory estimates employing  $V_n$  [...] are not able, by themselves, to produce estimates based on the objects' own motion. For example, that an object will fall, or that fast motion is more likely to be rectilinear than to traverse sharp angles over short intervals is not knowledge brought to the table by  $V_n$ . (Grush 2007b, pp. 406–407)

<sup>7</sup> Human mind is here understood as a controller of some (usually bodily) processes. This control loop is enormously complex and intricate, and in my opinion it is quite plausible to regard it within the framework of 4E cognition (Menary 2010), even if Grush's description and particular examples of this type of modelling might seem to accept some "skull-bound" account of mind.

<sup>8</sup> Grush notes a relevant difference between the *modal* and *amodal* emulation: the former is concerned directly with the incoming signals, while the latter with the basis functions (or some other transformations) of the input. This difference is crucial in the case of Otto the music genius—while he has learned the *modal* emulator, he has not learned the *amodal* emulator which is crucial for phenomenological aspects of having spatial purport (Grush 2007b, p. 406).

There are two matters in this quote. First of all, Grush points out that amodal emulation presupposed by the skill theory is insufficient to account for object movements in general. This is correct with regard to his model, as he does not introduce learned expectations about behavior of external objects into the emulator described by his theory. (There is, however, a slightly more convoluted possibility to do so by including some of the elements of the larger emulation framework. I will discuss this possibility in a moment.) Second of all, the author claims that this is not an issue, since he wants to give an account of perception of behavioral space which contains movement of objects only in so far, as it results from self-movement. There is no argument for this point in his paper, and this assumption is in my opinion an incorrect and implausible characterisation of behavioral space.

Whenever an object is thrown in our direction, and we can, even slightly and unawarely, perceive it, it prompts an action. Depending on the speed and exact trajectory of the movement we can try to get out of its way, catch it, deflect it, and sometimes even throw ourselves in its way, get hit on purpose, to protect somebody or something else.<sup>9</sup> These are full-fledged, intentional actions, not mere reactions. An object that is flying in our way prompts an opportunity to act, hence it becomes a part of our behavioral space. Note that this perception is as well imbued with spatial purport, as we perceive the directionality and distance of the moving objects in a dynamical fashion. This, I believe, shows that the issue of object-motions is not orthogonal to spatial purport and should be explicitly included in the skill theory.

What's surprising, is that Grush leaves his model like that, even though in his work on time he has referred to issues of object-movements, most notably the perception of apparent and biomechanical motions, which serve as the core presentation of the abilities of his trajectory estimation model (Grush 2005). How, then can we fill this gap, remaining within the emulation framework?<sup>10</sup>

The most conspicuous possibility is conceptually quite simple. We could include those motions by enhancing the environment or process model that is a part of the KF. However, the simplicity of this solution is misleading, as it amounts to introducing a black-box module defined as *some* function describing the evolution of the objects in our environment that is then moved outside of the immediate scope of the theory in question. (The general assumption that the modeled domain is driven by a Gauss-Markov process (Grush 2005, p. S211) does limit the choice of possible functions significantly, but is far from a sufficient exposition.) Then, this module is used for emulation and trajectory estimation. But not only we do not know what this function exactly does, Grush doesn't also account for its origin, i.e. how is it learned. Hence, this solution is unsatisfactory.

To offer a more detailed description of this black-box model of world evolution over time I will turn now to the insights provided by the predictive processing theory.

---

<sup>9</sup> Cf. a "Friends" episode "The One With The Ride-Along" where Joey throws himself in the way of a "bullet" to protect his meatball sub.

<sup>10</sup> I am thankful to an anonymous reviewer for pointing out that the TEM in general addresses this objection, even though Grush explicitly glosses over it in his presentation of the skill theory v2.0.

## 4.1 Insights from the predictive processing theory

Despite obvious similarities, hitherto only one attempt has been made to connect Grush's work within the emulation framework to PP. I am referring to Wanja Wiese's work on phenomenology of time perception and his HiTEM model (Wiese 2017), a hierarchical extension of Grush's trajectory estimation model (Grush 2006, 2009). I believe that such a straightforward move from the emulation framework to PP as Wiese does is, however, slightly problematic, and some more detailed discussion of the differences between those two views on cognition is required.

To make it as brief as possible: first of all, emulation framework is computationally "flat". Namely, even though Grush considers a system comprising of low-level modal emulators as well as of higher-level amodal emulators (Grush 2004, fig. 7 and text on p. 389), the emulators are imagined as working in parallel so that their combined estimates provide greater accuracy. Grush does not discuss any way of nesting one, lower-level emulator within another, higher-level one. This is, however, a minor issue, since "layering up" of emulators can be done quite easily and elegantly, as shown by Wiese's work (2017).

Second of all, in Grush's own words, his framework offers a way out of the top-down/bottom-up "dilemma": "Kalman gain allows us to breathe some much-needed flexibility and content into the stale and overly metaphorical distinction between top-down and bottom-up sensory/perceptual processing" (Grush 2004, p. 383). Hence, within KF-based processing scheme such a distinction is pointless. Both sensory data and emulation give rise to the conscious experience of the subject. However, PP is much more strongly related to this distinction, since there is a substantial difference between the error signal carried via the bottom-up pathways and complete percepts that inhabit the top-down streams.<sup>11</sup>

Let's unpack this idea a little bit: first of all, both in PP and within emulation framework the existence of the top-down pathway is a necessary condition for perception to occur. This is the point where those two ideas diverge from the so-called classic, bottom-up account of cognition. Second of all, in both cases the initial prediction is updated in accordance with the sensory data, that is weighed depending on its reliability (because of the precision estimates or Kalman gain, respectively). However, the account of bottom-up information processing provided by these two theories differs on the grounds of both model's structure and philosophical interpretation.

In case of PP, the use of information coming from the senses is indirect: we predict the world by predicting the probability distributions of sensations. Hence, when explaining away prediction errors, we compare them with the (actual) frequency distributions of stimuli (not stimuli themselves). This data is then used to statistically update the model (or for active inference) in order to offer a more accurate prediction. In case of emulation framework, the comparison of a posteriori prediction and actual sensory information is direct—Grush underlines that this is performed by the measurement inverse (see Fig. 1), and hence the residual correction is "translated"

<sup>11</sup> Here, following my own convictions, I will put this point in the radical wording even though the issue remains controversial and many supporters of the PP may disagree with such a position. To my knowledge, the most extensive discussion of this issue was offered by Paweł Gładziejewski (2017), and a recent paper by Andy Clark (2018b) offers more direct support for the radical position on the issue that is assumed here.

into the language of the states of physical world (more precisely, the language of the process model used by the filter). This constitutes a radical difference in the nature of content of the prediction error information and the residual correction. In other words, in case of Grush's framework, we compare two representations, identical in (semantic and metaphysical) nature, while in PP, we must find a "middle ground" for this comparison, by abstracting from the content of our representation, and focusing on its probabilistic properties.

Hence, to restate it in the philosophical language, it can be said that the emulation theory assigns some perceptual content or purport to the sensory data, while PP, as I understand it in this paper, claims that content can only originate in the internal model—since our conscious perceptual experiences, necessarily involving some intentionality, are direct results of its operations [such a view is maintained by Clark (2018b)]. This aspect of sensory inputs can be called their *contentlessness*, and by that I mean that the content of our conscious perceptual experience is in no direct way dependent on the senses. I do not mean here that the sensory signals are not about anything at all, that they are not intentional. This is a separate, fascinating issue that leads to the famous Sellarsian dilemma (see Gładziejewski 2017) and so—beyond the scope of this article. My claim here is only that even if the sensory deliveries are themselves intentional and contentful, their contents are not forwarded up the hierarchy, but translated into the prediction error signal which covers their probabilistic properties.<sup>12</sup>

## 4.2 Predictive and hierarchical skill theory

Including object-motions within the behavioral space has interesting consequences for the model of spatial purport. First of all, it nuances the role of PPC which was treated by Grush as a monolithic structure. In reality, it is a complicated region with an intricate internal organization, as exemplified by the hemispatial neglect, a syndrome resulting from a sustained brain injury or from lesions. Subject with hemineglect are characterized by reduced or lacking experience of the hemispace contralateral to the side of the damage. In result the patients have difficulties performing tasks that require intentional reaching or directing gaze towards the stimuli coming from the biased side (Heilman et al. 2000), while their performance of inattentive tasks, e.g. catching a ball surprisingly thrown at them, remains unchanged (Storey 2004). Hence it must be better understood how the sensory data are processed within the PPC and directly associated regions. One such area is the medial superior temporal area, which has cells specialized in perception of different aspects of movement and operates independently from attention (Milner and Goodale 2006, p. 48). MST is also immensely connected to regions outside of the so-called dorsal stream of visual processing (Milner and Goodale 2006), most importantly it has an independent processing stream (tectopulvinar route) going from the primary visual cortex via the pulvinar nucleus and middle temporal

<sup>12</sup> The point that I'm making here ties also into the discussion of the representational assumptions of predictive processing and reiterates Gładziejewski's observation, that while the generative model plays "a genuinely representational role, the same cannot be said about the sensory signal" (Gładziejewski 2017, p. 6; see also Gładziejewski 2016). Within emulation framework both sensory signals and emulators can be considered representational.

area. Finally, it is a relevant actor in the process of both self-motion and object-motion perception (Kleinschmidt 2002; Schenk and Zihl 1997a, b; Saygin 2007). A full discussion of the neuroscientific literature exceeds the scope of this paper, but recently a computational model based on Kalman filtering has been offered to account for the operation of the PPC with the areas directly connected to it in case of the motion induced position shift (MIPS) illusion (Kwon et al. 2015). MIPS can be observed when subjects are presented a stationary stimulus that exhibits some internal pattern motions (in case of Kwon, Tadin, and Knill's experiment simulated by Gaussian white noise). Then, they perceive the stimulus as shifted towards the direction of motion.

#### 4.2.1 Nested emulators

Kwon, Tadin, and Knill's model is a quite simple version of Kalman filter. It is defined with the following six parallel equations (Kwon et al. 2015, p. 2), fitted to explicitly account for the MIPS experiments conducted by the authors:

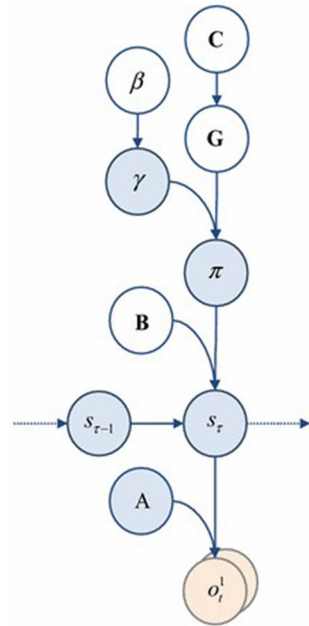
Model of motions in the world	Observation model
$x_t = x_{t-1} + \Delta t \cdot v_{t-1}^{obj}$ $v_t^{obj} = \alpha v_{t-1}^{obj} + \delta^{vo} \Omega_t^{vo}$ $v_t^{pattern} = \beta v_{t-1}^{pattern} + \delta^{vp} \Omega_t^{vp}$	$y_t^x = x_t + \eta^x \Omega_t^{y^x}$ $y_t^{vo} = v_t^{obj} + \eta^{vo} \Omega_t^{y^{vo}}$ $y_t^{vp} = v_t^{obj} + v_t^{pattern} + \eta^{vp} \Omega_t^{y^{vp}}$

The model of the world here (on the left of the table above) is the internal generative model that is used during perception, which itself is based on the observation model. The first term of the world model describes how the object's position (vector  $x_t$ ) at time  $t$  is influenced by its position ( $x_{t-1}$ ) and velocity ( $v_{t-1}^{obj}$ ) at time  $t - 1$  ( $\Delta t$  incorporates into the model the possibility of using different granulation of time, regarded here as discrete—as with any Kalman filter scheme it is possible to extend this into linear time, although discrete rendering of the model largely simplifies it and is sufficient for our current purposes). The other terms represent the stationary (and local) processes that the object undergoes (in Kwon, Tadin, and Knill's account those processes are *ex definitione* Gaussian):  $\alpha$  and  $\beta$  correlate the object and pattern velocities over time (this explains why zero object velocity is perceived as nonzero when the pattern velocity is nonzero—an aspect of human perception responsible for MIPS).  $\Omega_t$  accounts for unpredictable noise, in this specific case white Gaussian noise, while  $\delta$ s represent standard deviations of changes in object and object's pattern. Altogether  $\alpha$ ,  $\beta$ ,  $\delta$ , and  $\Omega$  modulate the weights of the incoming sensory signal (in terms of PP: resolve the amount of prediction error at time  $t$ ).

The observation model (on the right of the table above) offered by Kwon, Tadin, and Knill uses retinotopic coordinates of the object's position ( $y_t^x$ ), retinal velocity ( $y_t^{vo}$ ), and object pattern's retinal velocity ( $y_t^{vp}$ ). These are “corrupted” by the unpredictable (Gaussian) noise  $\Omega$ , with its standard deviation given by  $\eta$  (the sensory noise).

Those equations are the subject of the Kalman filter that, according to authors, calculates the specific values of  $v^{obj}$  and  $v^{pattern}$  from  $y$ —retinal velocity. Presented

**Fig. 2** A schematic presentation of a generative model directing active inference. See Kaplan and Friston (2018) for the description of variables  $A$ ,  $B$ ,  $C$ ,  $\pi$ ,  $o_t$ ,  $s_t$ .  $\gamma$  and  $\beta$  represent, respectively, precision of beliefs about policies and prior expectation of its inverse, while  $G$  stands for the expected free energy (explained further in the text). Adapted from (Kaplan and Friston 2018, p. 6), used on the CC BY 4.0 International license



this way, Kwon, Tadin, and Knill’s model offers a modal emulator. But we know from Pouget’s work that the posterior parietal cortex does not use directly the sensory inputs, but first calculates their basis functions. Hence, we may attach this Kalman filter to Grush’s skill theory, providing in this way an upper, hierarchical layer. The Kwon, Tadin, and Knill’s model can be construed as performing the role of a “second order” emulator, influencing the operation of the “first order” emulator, the Kalman filter postulated by Grush. They both use the same inputs (basis functions calculated from the sensory input), but the second order emulator does not directly affect the perception, but rather influences the work of the first order emulator, dynamizing it in a non subject-centric way, and enabling it to perceive object-motions (while remaining within the egocentric frame of reference).

This is the *hierarchical* part of the PHiST, although it will need to be slightly nuanced further in the text.

#### 4.2.2 Generative model

This, however does not suffice, as it does not cover the issues related to the direction of data processing. Such a layered up scheme is still framed in a way that presents this hierarchical skill theory as mainly bottom–up driven. To cope with this problem we must show how the model can be made more predictive with the *active inference framework* (Friston 2009, 2010; Friston et al. 2011; Friston and FitzGerald 2017).

Recently, within the active inference framework some work has been done to cover the issues related to space perception. Kaplan and Friston (2018) develop a model describing the performance of a navigation task, based on a generative model, depicted



in Fig. 2. Authors discuss a fairly simple environment, leading to an agent restricted to barely few policies (available strategies of acting) and sensory modalities. Even so, their model demands quite a bit of mathematical knowledge on the part of the reader. Because of that, for our current purposes I will attempt only to specify the interpretation of the variables the generative model is comprised of, without an enquiry into the mathematics behind them (this, as well as a computer simulation of the operation of this model, should be done during further research into the issues of spatial purport).

The easiest part to begin with are the notions of outcomes, outcome modalities, hidden states and policies. Instructed by Grush's work, we may reduce the outcome modalities to only one, since the PPC integrates data coming from multiple sensory modalities by calculating their basis functions, and the model, to remain amodal, will be operating on those data. This means that the outcome vector  $o_t$  (of shape  $n \times 1$ ) encodes the expected basis functions at time  $t$  (where  $n$  is the dimensionality of the basis functions). The hidden states correspond to the physical location of objects in subject's behavioral space. Finally policies are strategies of actions, overseeing the functioning of the agent over an extended period of time, hence they specify the driving force previously denoted as  $e(t)$  and action preparation and execution.

Model parameters are slightly more complex to explain. First of all, the likelihood  $A$  is (most likely) a multidimensional matrix. Since it calculates the conditional probability of an outcome given a hidden state ( $P(o_t|s_t)$ ), and there is only one outcome modality, it is responsible for the reduction of the (circa) three sensory modalities imbued with spatial significance by calculating the basis functions of their inputs. This means that neurally it is implemented within the PPC, and could be simulated according to the basis function model.

Now, the state transition probabilities matrix  $\mathbf{B}$  corresponds to the second-order emulator described in the previous section. This matrix describes the posterior expectation of probability of a given state  $\mathbf{s}_t$  arising from a previous state  $\mathbf{s}_{t-1}$ , accounting for the world's dynamics. Kaplan and Friston note also that this empirical prior transition matrix is prescribed by the policy the agent entertains at time  $t$ , so that the transition to the next state  $\mathbf{s}_{t+1}$  explicitly depends both on the previous state  $\mathbf{s}_t$  and on the policy  $\pi$ .

Finally, we need to define what prior beliefs  $C$  the agent holds. They are largely dependent on the goal the agent pursues at a given moment and its general knowledge of the environment dynamics, and since we want to discuss a more general ability than Kaplan and Friston did, it will be best to first offer some examples. In a simplistic case, similar to the  $8 \times 8$  grid labyrinth from the original paper, where an agent is personally located within a maze (not just looking at it), and for whatever reason imbued with a task to reach a given location, it will be likely to have the prior beliefs as Kaplan and Friston's model, namely to find itself in a location that is best accessible from the target location. If, located in the same maze, an agent will want to reach the exit, it will be likely to hold a belief of being in a location which is best accessible from a location that is not a part of the labyrinth. If the agent is hungry, the higher-order prior (i.e. a more general belief, conditioning the moment-to-moment operations of the generative model) of not being hungry (which follows from the will of maintaining agent's integral structure—necessary for long-term minimization of free energy) will result in a prior belief of being in a location where food is best accessible. Finally: if

an object gets thrown in my direction and I will predict that it will hit me soon, I will entertain a belief of not being hit (that again follows from the more general belief of maintaining integrity) and act accordingly (ie. assume proper policies)—either move to a location where I will not be hit, deflect the object, or hide myself from it.

This generative model complicates the Kalman-filtering scheme discussed hitherto. Most importantly, it truncates the processing that was earlier continuous, and changes slightly the language used by Grush (as e.g. in case of the likelihood matrix  $A$ , which roughly substitutes the measurement matrix  $O$  and its inverse  $O^{-1}$ ). But the introduction of the idea of prior beliefs  $C_t$  and the notion of policies  $\pi$ , intricately tied to the state transition matrix  $B$ , shows us how, in the predictive processing framework, online and offline spatial processing is related and co-dependent.

## 5 A model of spatial purport

It seems that the insights from the predictive processing framework provide the necessary tools to enhance Grush's skill theory and open the black box of environment model. Before I will show what does it do to our understanding of what spatial purport is, I will summarize the work done in the previous section and show how PHiST looks on the algorithmic level (in the sense of Marr 1982).

### 5.1 PHiST

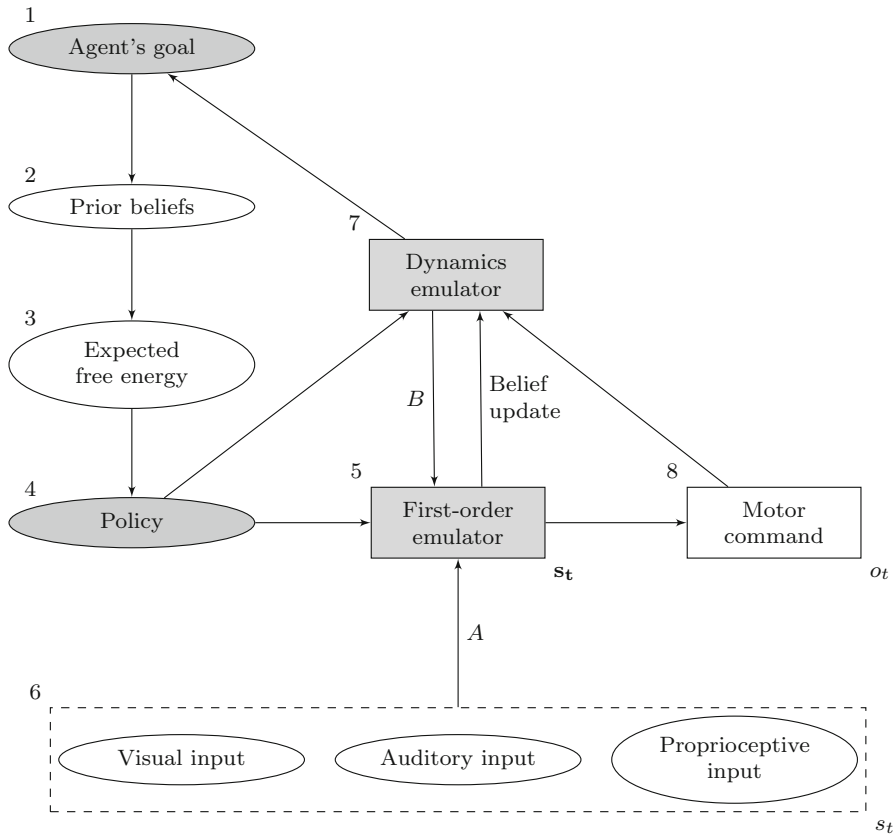
Figure 3 shows a diagram which abstracts from the formal and mathematical underpinnings of the model and instead attempts to show the general algorithm responsible for the perception of behavioral space (object motions included). The remainder of this section describes this diagram in more details.

At the top of the hierarchy (box 1) are agent's goals that define the task that the agent attempts to perform, which depend on more general properties of the agent, such as e.g. its homeostatic drive. The goal defines agent's prior beliefs (box 2) which are here regarded as task-specific. Their function is to reformulate the goal into preferences over outcomes.

Expected free energy (box 3) depends upon the context and the preferences over outcomes and plays an important role in specifying the policies the agent will entertain: by giving higher prior probability to those that minimize the value of expected free energy<sup>13</sup>.

Policies (box 4) are, as already discussed in Sect. 4.2, the schemes of actions that the agent will employ. Their role is to specify the motor commands necessary for calculation of the basis functions of sensory inputs and for the operation of the first order emulator (box 5), which is more or less the Kalman filter from original skill

<sup>13</sup> Recent results of Parr and Friston (2019) indicate that this step is unnecessary, as the prior beliefs could be incorporated into agent's generative model or, in our context—incorporated into agent's goal which, then, could specify the policies directly. However, given that the expected free energy is a specific case of generalized free energy and that this notion has a history within the predictive processing framework, I will retain steps depicted in boxes 2 and 3 for the sake of clarity.



**Fig. 3** A diagram of the PHiST, pitched at a high level of abstraction. Gray boxes cover those parts of the model that were in some form present in Grush's skill theory v2.0, while white boxes depict modules specific for the predictive processing framework. Labels near the top left corner of each box correspond to the description in the main text

theory v2.0, conveying the information that in Grush's model was provided by the driving force.

The most important departure from the original formulation regards the procedure for dealing with sensory data. In PHiST a twofold processing scheme is implemented as the belief update. First, the free energy is minimized internally, as gradient descent over predicted hidden states. Second, the process is outsourced to the dynamics emulator (second-order emulator, box 7), where it is used to update the predicted transition matrix  $B$ . The result of the computation is the motor command (box 8). In neural terms the main area involved is the PPC, as this is an amodal emulator operating on the basis functions of input. Basis functions are calculated from the sensory inputs (here from the three modalities identified as crucial for spatial purport) with the transformation matrix  $A$  (box 6) and are used for calculating the prediction errors between the incoming signals and predicted hidden states (real locations) of the world.

However, the first-order emulator is as well dependent on the dynamics emulator (box 7), the main novel element introduced by PHiST. The data it receives from the first-order emulator are used to maintain an estimate of the dynamics of the world (via the process of free energy minimization) which is represented by the matrix  $B$  [more precisely, by the posterior approximation  $\mathbf{B}$  (see Friston and FitzGerald 2017, pp. 13–14)] that is feeded back to the first-order emulator where it plays the role of Grush’s process model. Using the efferent copy of the motor command it is able to model motions that result from agent’s own actions. But by extracting the general statistical properties of received inputs, this model is able to also account for object motions both in a context-dependent (by learning properties of temporal successions of states present in its immediate surroundings) and context-independent (by maintaining some form of “intuitive physics”) way. An efferent copy from this second-order emulator is also passed back to the agent’s goal, enabling performance supervision.

Finally, the motor command (box 8) is the outcome of the model. Neurally speaking, this is the motor command feeded by the policy (box 4) with the exact spatial coordinates specified by the first-order emulator (box 5) together with the dynamics emulator (box 7). As a result of the operation of this processing scheme, the agent is able to take skillful actions in its environment.

While this proposed model specifies the general process theory of perception of behavioral space, future research will have to provide precise, mathematical formulations of the operations described above, as well as test the plausibility of this model. Hopefully, however, the above specification is precise enough to serve as a bedrock for empirical and simulation studies. What remains, however, is to see how this proposed model ties back to the discussions of spatial purport.

## 5.2 Understanding spatial purport of perceptual experiences

The question remains, however, how the proposed model accounts for spatial purport (or content, or phenomenology).

In Sect. 2.3 I have identified the experiential loop that ties together the experiences of distance and directionality with the experiences of self-location, and grounds them all in dispositions to act or skills. Predictive and Hierarchical Skill Theory, as outlined in the previous section, if it is a correct model of spatial purport, implements this dependence. The model’s outcomes are motor commands populated with precise spatial (or spatio-temporal) coordinates, that, when enacted, should result in skillful actions in the environment. What’s more, not only spatial purport is geared towards behavior, but it’s also dependent on it, as the operation of the two emulators necessarily depends on agent’s policies.

PHiST includes also the multimodal nature of spatial purport, suggesting another argument for the Self-Location Thesis discussed in Sect. 2.2.1. Vision in the course of evolution has usually come associated by other sensory modalities, and as I have argued previously, it doesn’t seem that anyone would hold that multimodal sensory experience does not involve location of the perceiver among its face value contents. So, even if in virtue of some of its essential properties visual experience does not involve self-location data, vision as it is realised in the real world does, due to its inter-

connectedness with other senses. Hence, if we consider an unusual case in which only information delivered by eyes is in question, it can still provide self-locating contents as the information is then processed by the very same mechanism as it would normally be processed together with other senses, which imbues the experience with spatial purport.

Note that this doesn't conflate the initial distinction between an experience carrying spatial information and one imbued with spatial purport. This is the case when a sensory modality which normally does not provide sensory information starts to. Consider, once again, the exemplary case of Inga and Otto from the introduction. Informed by PHiST, how can we describe what happens in Inga's cognition that is not shared by Otto? Obviously, they have largely different generative models of the world: most importantly Otto's PPC does not process the auditory inputs (or at least some of its properties, such as pitch) he receives from the sensory augmentation device, what means that his *likelihood mapping A* contains no description of the relation between such auditory hidden states and outcomes. This is more or less the same statement as Grush's claims that the spatial purport appears only if one's emulator is amodal, but now it is also clear how come if Otto were to learn the modal emulator perfectly, he still would not have any spatial content. Learning a modal emulator would lead him to creating a new generative model (or simply would update some other, unrelated model), one using entirely different policies and state transition matrices—different not only with regard to their “content”, but most importantly to the representations they operate on. Such a model could then be, obviously, *calibrated*, as Grush (2007a) describes the process of learning the coordination between “quasi-spatial” and “genuinely spatial” experiential manifolds, but this process would not imbue the new generative model with spatial purport (or at least with the same spatial purport as the original model). It would still represent entirely different policies and transition matrices and, probably, hidden states, outcomes, and likelihood mappings as well, i.e. it would be sensitive to different (or differently construed) aspects of the environment.

The discussion of spatial purport nicely ties back to the PP story of momentary perceptual experience of the world at large, briefly mentioned earlier. PP claims, that

momentary perceptual experience [...] always reflects a delicate combination of top-down model-based prediction, self-estimated sensory uncertainty, and bottom-up (incoming) sensory evidence. When the top-down prediction is wrong, “prediction error” signals result, and these are fed upwards (and sideways), allowing new “top-down” guesses rapidly to be recruited. It is only when all that settles, within current tolerances of noise, that a clear percept is formed. (Clark 2018b, p. 72)

As such, momentary percepts are strongly related to action—this is what makes them “Unitary-Coherent”, rather than poised with a “Bayesian blur” of possibilities” (Clark 2018b). If this is truly the nature of perceptual experience, spatial purport may have slightly different, more foundational role, than initially thought. As we see from PHiST, spatial purport disambiguates action, providing the general action-type specified by the policy with information necessary to perform a particular action-token. Hence, spatial purport underlies all action and in so far as action underlies all perceptual experiences, spatial purport underlies it as well.

In this sense, spatial purport permeates a majority of our experiences, constituting a basic aspect of our perspective on the world. But it is neither necessary nor trivial—it should be obvious from the case of Inga and Otto, as Otto’s experience becomes (artificially) devoid of spatial content. If at the same time he is located in an unknown, or quickly changing environment, he becomes completely unable to do anything other than taking up the process of blindfold hypothesis–testing by making minimal moves and testing how far he can go, at the same time not risking too much from his current, apparently safe position (hence showing the interplay of novelty seeking and conservatism permeating human behavior, known also under the name of the “darkened room puzzle”, see e.g. Friston et al. 2012; Clark 2018a). However, if Otto would be in this moment in his own apartment it is extremely likely that the repertoire of actions available to him would be much larger, and he could move around with greater certainty. This shows that the spatial phenomenology is not at all in the sensory sensations, but instead in the generative model the cognitive system possesses.

## 6 Summary

To sum up, despite several issues, Gareth Evans’ and Rick Grush’s accounts of how the spatial purport arises in result of the subject’s (bodily) actions in their environment offer strong foundations for proposing a mechanism underlying the specific, phenomenal aspects of spatial perception that Grush termed “spatial purport”. After working out several incompatibilities, I was able to devise a new theory that takes what’s best in Grush’s explanation and combines it with the insights from the predictive processing theory and active inference framework. This account, Predictive and Hierarchical Skill Theory, offers a redescription of the explanandum of the theory, hopefully opening up possibilities of empirical examination of both original skill theory and PHiST.

This account is, I believe, interesting not only to a cognitive scientist, but also to the philosophers of mind, since working out how the spatial purport arises enabled us to offer a new, more concise description of what it actually *is*. Following Evans and Grush, I showed in the last section that spatial purport is closely tied to subject’s ability to actively participate in their environment. But the proposed account assigns to spatial purport a significant, foundational, and so far overlooked role in our experiential lives that requires further examination.

This conclusion resonates with the embodied and enactive perspective on perception and cognition. I believe that spatial purport opens up a possibility of seeing that these two traditions that PP continues have a strong common core that can be uncovered through further research, leading to a deeper understanding of the nature of the mind.

**Acknowledgements** I want to thank Paweł Gładziejewski, Joanna Komorowska-Mach, Marcin Miłkowski, Jelle Bruineberg, and an anonymous reviewer for helpful feedback on earlier versions of this manuscript that has enabled me to greatly improve the paper. This work was funded by the Ministry of Science and Higher Education (Poland) research Grant DI2018 010448 as part of “Diamentowy Grant” program.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Andersen, R. A., & Buneo, C. A. (2003). Sensorimotor integration in posterior parietal cortex. *Advances in Neurology*, *93*, 159177.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577–609. (discussion 610–660).
- Buneo, C. A., & Andersen, R. A. (2006). The posterior parietal cortex: Sensorimotor interface for the planning and online control of visually guided movements. *Neuropsychologia*, *44*(13), 2594–2606. <https://doi.org/10.1016/j.neuropsychologia.2005.10.011>.
- Cassam, Q. (2005). *Self and world*. Oxford: Oxford University Press.
- Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford: Oxford University Press.
- Clark, A. (2018a). A nice surprise? Predictive processing and the active pursuit of novelty. *Phenomenology and the Cognitive Sciences*, *17*(3), 521534. <https://doi.org/10.1007/s11097-017-9525-z>.
- Clark, A. (2018b). Beyond the 'Bayesian Blur': Predictive processing and the nature of subjective experience. *Journal of Consciousness Studies*, *25*(3), 71–87.
- Grush, R. (2007a). Berkeley and the spatiality of vision. *Journal of the History of Philosophy*, *45*(3), 413–442. <https://doi.org/10.1353/hph.2007.0060>.
- Grush, R. (2007b). Skill theory v2.0: Dispositions, emulation, and spatial perception. *Synthese*, *159*(3), 389–416. <https://doi.org/10.1007/s11229-007-9236-z>.
- Grush, R. (2009). Space, time, and objects. In J. Bickle (Ed.), *The Oxford handbook of philosophy and neuroscience*. New York: Oxford University Press.
- Evans, G. (1982). *The varieties of reference*. (J. H. McDowell, Ed.) New York: Clarendon Press.
- Evans, G. (1985). "Molyneux's question". In *Collected papers*. Oxford: Clarendon Press. ISBN: 978-0-19-824737-1.
- Friston, K. J. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293–301. <https://doi.org/10.1016/j.tics.2009.04.005>.
- Friston, K. J. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. <https://doi.org/10.1038/nrn2787>.
- Friston, K. J., FitzGerald, T., et al. (2017). Active inference: A process theory. *Neural Computation*, *29*(1), 1–49. [https://doi.org/10.1162/NECO\\_a\\_00912](https://doi.org/10.1162/NECO_a_00912).
- Friston, K. J., Mattout, J., & Kilner, J. (2011). Action understanding and active inference. *Biological Cybernetics*, *104*(1), 137160. <https://doi.org/10.1007/s00422-011-0424-z>.
- Friston, K. J., Thornton, C., & Clark, A. (2012). Free-energy minimization and the dark-room problem. *Frontiers in Psychology*, <https://doi.org/10.3389/fpsyg.2012.00130>.
- Gharbawie, O. A., Stepniewska, I., & Kaas, J. H. (2011). Cortical connections of functional zones in posterior parietal cortex and frontal cortex motor regions in new world monkeys. *Cerebral Cortex*, *21*(9), 1981–2002. <https://doi.org/10.1093/cercor/bhq260>.
- Gładziejewski, P. (2016). Predictive coding and representationalism. *Synthese*, *193*(2), 559–582. <https://doi.org/10.1007/s11229-015-0762-9>.
- Gładziejewski, P. (2017). The evidence of the senses. In T. K. Metzinger & W. Wiese (Eds.), *Philosophy and predictive processing*. Frankfurt am main: MIND Group.
- Grush, R. (1998). Skill and spatial content. *Electronic Journal of Analytic Philosophy* *6*(6). Retrieved from <http://ejap.louisiana.edu/EJAP/1998/grusharticle98.html>. Accessed 26 May 2020.
- Grush, R. (2000). Self, world and space: The meaning and mechanisms of ego-and allocentric spatial representation. *Brain and Mind*, *1*(1), 5992. <https://doi.org/10.1023/A:1010039705798>.



- Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences*, 27(3), 377–396. (discussion 396–442).
- Grush, R. (2005). Internal models and the construction of time: Generalizing from state estimation to trajectory estimation to address temporal features of perception, including temporal illusions. *Journal of Neural Engineering*, 2(3), S209–S218. <https://doi.org/10.1088/1741-2560/2/3/S05>.
- Grush, R. (2006). How to, and how not to, bridge computational cognitive neuroscience and Husserlian phenomenology of time consciousness. *Synthese*, 153(3), 417–450. <https://doi.org/10.1007/s11229-006-9100-6>.
- Hafting, T., et al. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, 436(7052), 801–806. <https://doi.org/10.1038/nature03721>.
- Heilman, K. M., Valenstein, E., & Watson, R. T. (2000). Neglect and related disorders. *Seminars in Neurology*, 20(4), 463–470. <https://doi.org/10.1055/s-2000-13179>.
- Kaplan, R., & Friston, K. J. (2018). Planning and navigation as active inference. *Biological Cybernetics*, 112, 323–343. <https://doi.org/10.1007/s00422-018-0753-2>.
- Kleinschmidt, A. (2002). Neural correlates of visual-motion perception as object- or self-motion. *NeuroImage*, 16(4), 873–882. <https://doi.org/10.1006/nimg.2002.1181>.
- Konig, S. U., et al. (2016). Learning new sensorimotor contingencies: Effects of long-term use of sensory augmentation on the brain and conscious perception. *PLoS ONE*, 11(12), e0166647. <https://doi.org/10.1371/journal.pone.0166647>.
- Kwon, O.-S., Tadin, D., & Knill, D. C. (2015). Unifying account of visual motion and position perception. *Proceedings of the National Academy of Sciences*, 112(26), 8142–8147. <https://doi.org/10.1073/pnas.1500361112>.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. Cambridge, MA: MIT Press.
- Menary, R. (2010). Introduction to the special issue on 4E cognition. *Phenomenology and the Cognitive Sciences*, 9(4), 459–463. <https://doi.org/10.1007/s11097-010-9187-6>.
- Metzinger, T. K. (2004). *Being no one: The self-model theory of subjectivity*. Cambridge, MA: MIT Press.
- Milner, A. D., & Goodale, M. A. (2006). *The visual brain in action*. Oxford psychology series (2nd ed., Vol. 43). Oxford: Oxford University Press.
- Moser, E. I., Kropff, E., & Moser, M.-B. (2008). Place cells, grid cells, and the brain's spatial representation system. *Annual Review of Neuroscience*, 31(1), 69–89. <https://doi.org/10.1146/annurev.neuro.31.061307.090723>.
- O'Keefe, J., et al. (1998). Place cells, navigational accuracy, and the human hippocampus. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 353(1373), 1333–1340. <https://doi.org/10.1098/rstb.1998.0287>.
- O'Regan, J. K., & Noë, A. (2001). What it is like to see: A sensorimotor theory of perceptual experience. *Synthese*, 129(1), 79–103. <https://doi.org/10.1023/A:1012699224677>.
- Parr, T., & Friston, K. J. (2019). Generalised free energy and active inference. *Biological Cybernetics*, 113(5), 495–513. <https://doi.org/10.1007/s00422-019-00805-w>.
- Pouget, A., Deneve, S., & Duhamel, J.-R. (2002). A computational perspective on the neural basis of multisensory spatial representations. *Nature Reviews Neuroscience*, 3(9), 741–747. <https://doi.org/10.1038/nrn914>.
- Pouget, A., & Sejnowski, T. J. (1997). Spatial transformations in the parietal cortex using basis functions. *Journal of Cognitive Neuroscience*, 9(2), 222–237. <https://doi.org/10.1162/jocn.1997.9.2.222>.
- Saygin, A. P. (2007). Superior temporal and premotor brain areas necessary for biological motion perception. *Brain*, 130(9), 2452–2461. <https://doi.org/10.1093/brain/awm162>.
- Schenk, T., & Zihl, J. (1997a). Visual motion perception after brain damage: I. Deficits in global motion perception. *Neuropsychologia*, 35(9), 1289–1297. [https://doi.org/10.1016/S0028-3932\(97\)00004-3](https://doi.org/10.1016/S0028-3932(97)00004-3).
- Schenk, T., & Zihl, J. (1997b). Visual motion perception after brain damage: II. Deficits in form-from-motion perception. *Neuropsychologia*, 35(9), 1299–1310. [https://doi.org/10.1016/S0028-3932\(97\)00005-5](https://doi.org/10.1016/S0028-3932(97)00005-5).
- Schwenkler, J. (2014). Vision, self-location, and the phenomenology of the 'point of view': Vision, self-location, and the phenomenology of the 'point of view'. *Noûs*, 48(1), 137–155. <https://doi.org/10.1111/j.1468-0068.2012.00871.x>.
- Seth, A. K. (2014). A predictive processing theory of sensorimotor contingencies: Explaining the puzzle of perceptual presence and its absence in synesthesia. *Cognitive Neuroscience*, 5(2), 97–118. <https://doi.org/10.1080/17588928.2013.877880>.

- Storey, E. (2004). The neglect syndrome. In *Neurology and clinical neuroscience* (pp. 73–82). Elsevier. <https://doi.org/10.1016/B978-0-323-03354-1.50010-9>. ISBN: 978-0-323-03354-1
- Wiese, W. (2017). Predictive processing and the phenomenology of time consciousness. In T. K. Metzinger & W. Wiese (Eds.), *Predictive processing and philosophy*. Mainz: Theoretical Philosophy/MIND Group-JGU Mainz. <https://doi.org/10.15502/9783958573277>.
- Wiese, W., & Metzinger, T. K. (2017). Vanilla PP for philosophers: A primer on predictive processing. In T. K. Metzinger & W. Wiese (Eds.), *Philosophy and predictive processing*. Frankfurt am main: MIND Group.
- Zarchan, P., & Musoff, H. (2000). *Fundamentals of Kalman filtering: A practical approach* (Vol. 190)., Progress in astronautics and aeronautics Reston, VA: American Institute of Aeronautics and Astronautics.
- Zipser, D., & Andersen, R. A. (1988). A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature*, 331(6158), 679–684. <https://doi.org/10.1038/331679a0>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.