

Skill theory v2.0: dispositions, emulation, and spatial perception

Rick Grush

© Springer Science+Business Media B.V. 2007

Abstract An attempt is made to defend a general approach to the spatial content of perception, an approach according to which perception is imbued with spatial content in virtue of certain kinds of connections between perceiving organism's sensory input and its behavioral output. The most important aspect of the defense involves clearly distinguishing two kinds of perceptuo-behavioral skills—the formation of dispositions, and a capacity for emulation. The former, the formation of dispositions, is argued to be the central pivot of spatial content. I provide a neural information processing interpretation of what these dispositions amount to, and describe how dispositions, so understood, are an obvious implementation of Gareth Evans' proposal on the topic. Furthermore, I describe what sorts of contribution are made by emulation mechanisms, and I also describe exactly how the emulation framework differs from similar but distinct notions with which it is often unhelpfully confused, such as sensorimotor contingencies and forward models.

Keywords Spatial perception · Skill theory · Sensorimotor contingencies · Emulation theory

1 Introduction

The issue of the spatial content of perception has a long history. Nearly as long has been the attempt to connect, in one way or another, the spatial content of perception to actual or possible behavior. This approach has seen something of a flurry of interest lately (Mel 1986; Grush 1995; Grush 1998; Noë 2004, 2006). But while I

R. Grush (✉)
Department of Philosophy, UC San Diego, 9500 Gilman Drive,
La Jolla, CA 92093-0119, USA
e-mail: rick@mind.ucsd.edu

think that these accounts have been on the right track to one degree or another, they (mine included) have also been underspecified and confused. In this paper I will present a schematic version of what I currently believe to be a suitably specified and unconfused version of the theory. It is schematic in that for a full explication of some of the components of the theory, I will need to refer the reader to other works, specifically Grush (1998, 2000, 2004a,b, 2005, 2007b). I will attempt to provide in this paper glosses of these components sufficient to afford a good intuitive idea of the theory. But I am keen to be clear that the full theory is distributed across these sources, and the specialist interested in the detailed version should consult these sources together. I'm aware of the fact that this sort of circumstance ought to be addressed by a book. I'm working on that. But for now, to the theory.

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 and spatial relations (in a sense I will discuss later), 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' (which I used in an extended discussion of Berkeley's (1948) views on the spatiality of vision in Grush 2007a) 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.

There are two related kinds of spatial representation I will discuss: *egocentric location* and *shape*. I will not say anything directly about shape until Sect. 4.3. Egocentric spatial representation is typically distinguished from objective spatial representation. The rough distinction is clear enough: my representation of Oregon as lying between California and Washington in no way depends on *my* spatial location or relations, and so it is objective; while my visual experience of the coffee cup represents it as *just ahead to the left*, and the content of representation does make crucial reference to me. There are three relevant aspects of the space in terms of which my perceptual experience represents the cup. The first is the *origin*, which is roughly my body, less roughly it is perhaps my eyes or torso. The second is the *direction*: 'to the left' doesn't mean 'East' or 'West' or 'uphill.' Rather, its meaning derives from my body, particularly behavioral capacities and axial asymmetries. 'To the left' in this context just means something like 'what I could look at by turning my head thusly' or 'what I could point to (or grasp) by moving my arm and hand like such-and-so.' Third and finally, the *magnitudes* are not in objective units. While my experience presents the cup as very precisely localized—precise enough that I can quickly and easily grasp it, getting my fingers within millimeters of the surface in a fraction of a second—I could only make the roughest guess at its distance in centimeters or inches. So the magnitudes are also given in behavioral terms. Because the features of this space are defined in behavioral terms, I will follow Gareth Evans in calling this space the *behavioral space*. It is the space that is made manifest in perceptual experience, the space defined by the organism's body and possibilities for behavior, as shall be made significantly more

precise below. The expression ‘egocentric’ suggests that the space is centered on the ego, but it is silent on the nature of directions and magnitudes; so that in some strict etymological sense of ‘egocentric,’ representing something as being 3 meters north of me would qualify as an *egocentric spatial* representation. It would not qualify as a *behavioral spatial* representation, though.

So from now on I will drop ‘representation,’ ‘content,’ ‘egocentric,’ and other terms and speak of *behavioral spatial purport*. To further clarify what I mean by this expression it will help to look at the difference between cases in which there is, and is not, such purport. My example will concern the *sonic guide*, a member of the family of sensory substitution devices whose main purpose is to provide distance senses for the blind (my discussion of the sonic guide is quite abbreviated—see Grush 1998, 2000 for more detailed discussion). The sonic guide consists of a device worn on the head that includes a transmitter that emits a continuous ultrasonic probe tone, and two microphones that pick up reflections of that probe tone. The microphone’s signals are translated into audible sound profiles that are presented to the subject through earphones—for example, echoes from distant objects are translated into high-pitched sounds, weak echoes into lower volumes.

As Heil (1987) describes it, the

...sonic guide taps a wealth of auditory information ordinarily unavailable to human beings, information that overlaps in interesting ways with that afforded by vision. Spatial relationships, motions, shapes, and sizes of objects at a distance from the observer are detectable, in the usual case, only visually. The sonic guide provides a systematic and reliable means of hearing such things.

I will suppose what seems to be plausible, that subjects who have been using the device for a while and are competent with it are actually *perceiving* the objects in their environment directly, rather than *reasoning out* what the environment must be like on the basis of pitches and volumes (see Bower 1977; Aitken and Bower 1982). This seems to be accepted by Heil who, in discussing the sonic guide notes:

Devices like the sonic guide...prove useful only after the sensations they produce become transparent. ...successful use of the sonic guide requires one to hear things and goings-on rather than the echoes produced by the device. ... [children] less than about 13 months... do this quite naturally, while older children do so only with difficulty.

Now for a thought experiment. Consider a subject, call her ‘Toni,’ who is congenitally blind but who has been wearing the guide from birth. And let us assume that through the guide Toni has rich and direct perceptual experience of her immediate environment—comparably rich and direct as the experience normal subjects enjoy through their eyes. Compare Toni’s experience to the experience I would have if I were to don the guide. For me, the deliverances of the guide will be nothing but a strange cacophony. One small element of the highly variable cacophony might be a tone sounding at a pitch of Middle C at a volume of 35 dB. For me, this elicits experience *of a tone at a certain pitch and volume*. For Toni it elicits experience *of an*

*object just ahead there on the ground.*¹ Same guide, same sensory signal, very different experience, very different phenomenology. This difference between Toni and me is exactly the difference that I intend to capture with the expression *purport*. Toni's experience has behavioral-spatial purport, mine does not.

In what, exactly, does this difference consist? It does not consist in any difference in the presented auditory signal, nor in the information carried by that signal (the signal has features, discriminable by both Toni and myself, that covary with location in behavioral space). Second, it need not consist in any different capacities on the part of Toni or myself to discriminate relevant features (pitch, volume) of the signal—I may be able to tell with great accuracy and precision the signal's pitch and volume. So these are some things in which the difference does not consist. In what does it consist? Here are some natural suggestions: Toni is used to the guide, and I am not; for Toni, the guide causes experience with behavioral-spatial phenomenology, and for me it does not; Toni, because of her experience with the guide, is able to quickly and automatically exploit the audible information to guide her actions. These are right, but they lack precision. Adding precision is one of the things I hope to accomplish.

I will close this introduction with a brief outline of the subsequent sections. In Sect. 2 I will give a very brief synopsis of what I am now inclined to call Gareth Evans' *disposition theory* of behavioral spatial purport, and then I will discuss the neural information processing mechanisms that underlie (most of) Evans' disposition theory, the *basis function model*. In Sect. 3 I do a number of things. First, I describe the emulation theory of representation, emphasizing four features: application to motor control, vision, the kind of knowledge employed in the process model, and the distinction between modal and amodal emulation. Second, I show how to combine the emulation theory with the basis function model. With the materials of Sects. 2 and 3 in hand, Sect. 4 is where I provide a step-by-step explication of a theory of the behavioral-spatial purport of perception. I also discuss application to shape. Section 5 is a general discussion in which, among other things, I briefly compare the resulting account to related accounts.

2 Disposition theory and the basis function model

I turn now to Evans' theory of behavioral spatial purport. In discussing the example of how something can be heard to be in some direction, Evans writes:

The subject hears the sound as coming from such and such a position, but how is this position to be specified? We envisage specifications like this: he hears the sound *up*, or *down*, *to the right* or *to the left*, *in front* or *behind*, or *over there*. It is clear that these terms are *egocentric* terms: they involve the specification of the position of the sound in relation to the observer's own body. But these egocentric terms derive their meaning from their (complicated) connections with the actions of the subject... (Evans 1985, p. 384)

¹ The skill theory that I will be articulating is meant to address the spatial content of perception, not other aspects. And so the account is meant to address the 'just ahead there' part of the perceptual content, not the 'object' part.

Auditory input, or rather the complex property of auditory input which codes the direction of the sound, acquires a spatial *content* for an organism by being linked with behavioral output... (Evans 1985, p. 385)

And in discussing the example of a blind subject's tactile exploration of an object such as a chair, Evans writes:

... when he uses his hand, the blind man gains information whose content is partly determined by the dispositions he has thereby exercised—for example, that if he moves his hand forward such-and-such a distance and to the right he will encounter the top part of the chair. And when we think of a blind man synthesizing the information he receives by a sequence of haptic perceptions of a chair into a unitary representation, we can think of him ending the process by being in a complex informational state which embodies information concerning the egocentric location of each of the parts of the chair; the top over there to the right (here, he is inclined to point or reach out), the back running from here to there, and so on. Each bit of information is directly manifestible in his behavior... (Evans 1985, p. 389)

...we must say that having the perceptual information at least partly consists in being disposed to do various things.... (Evans 1985, p. 383)

The common theme presented in these quotes is a connection between sensory input and behavioral dispositions. One could say, in summarizing Evans' position, that a sensory input comes to be imbued with behavioral-spatial purport for an organism to the extent that that input induces dispositions for spatial behavior.

It will be useful to get clear on what is meant by 'disposition' in this context. There is a crucial distinction, not made explicitly by Evans, between what I shall call type-selecting dispositions and detail-specifying dispositions. The disposition theory hinges on the latter. A type-selecting disposition is something about the stimulus that motivates the execution of this or that behavior type, as opposed to nothing or some other behavior type. For instance, a bright flash might motivate a head turn and foveation, but not a grasp; an itch might motivate an arm and hand movement and scratch, but not any eye movement. A detail-specifying disposition is a disposition that, for any given behavior type (such as a grasp or foveation, or whatever), specifies the details of how that behavior type will be executed if it is executed. So for example will my intended grasp (behavior type) be implemented by moving my hand like *this*, or like *that*? Of course, situations that elicit type-specifying dispositions often also simultaneously elicit detail-specifying dispositions. The flash elicits a type-selecting disposition to foveate, and also elicits the detail-specifying dispositions that guide that foveation. But detail-specifying dispositions are often elicited without a concomitant type-selecting disposition. Many of the objects in my visual field are such that I am not inclined to direct behavior of any type at them, but are also clearly inducing detail-specifying dispositions, in that there is no question how I would move my eyes or hands if I were to execute one of those behavior types directed at one of those objects.

What is relevant for the disposition theory, as I am interpreting it, is the detail-specifying disposition. Whether a stimulus motivates looking vs. grasping vs. scratching vs. fleeing, or nothing at all, is not directly relevant to the behavioral spatial purport induced by that stimulus. What is relevant are the detail-specifying dispositions that are induced. If, supposing the behavior type in question is a grasp, am I disposed to reach to the left or to the right?

In [Grush \(1998\)](#) I coined the expression ‘skill theory’ as a label for my attempt to explicate and defend Evans’ views—I was partially inspired by [Cussin’s \(1992\)](#) heavy use of the notion of skill in his discussions of Evans. Despite the fact that this name has propagated to some extent through the literature, I now think that it is misleading since it unhelpfully lumps together dispositions and various kinds of skills (see [Sects. 4 and 5](#) for the distinctions). For now, I will stipulate the name *disposition theory* for the theory of a specific kind of behavioral spatial purport, as described above and elaborated more below. There are kinds of skills and behavior-manifested knowledge that I will also discuss, in later sections, that are not dispositions, and these will account for different aspects of spatial purport. Rather than discuss Evans’ views in more depth (I direct the interested reader to [Grush 1998](#) for this), I will turn to a discussion of the neural information processing mechanisms that, on the hypothesis I am pushing, underlie behavioral spatial purport, and the result will be an extremely straight-forward and compelling interpretation and vindication of the disposition theory.

The posterior parietal cortex (henceforth PPC) is arguably the most important cortical area for representing egocentric space ([Buneo and Andersen 2006](#)). But surprisingly there is nothing in this area that resembles a topographic map. Single cell recordings—the usual tool for finding topographic representations—have failed to even hint at anything resembling such a map in the PPC. What has been found are cells that respond to combinations of sensory and postural signals. Sensory signals include things like signals about what is projecting onto the retinae, and where on the retinae it is projecting. Postural signals will include information about how the eyes are oriented in the head, or how the neck or legs are comported with respect to the torso, and so forth.

Following the work of [Zipser and Andersen \(1988\)](#) and [Pouget \(Pouget and Sejnowski 1997; Pouget et al. 2002\)](#), we can describe the way that these sensory and postural signals are combined as follows. For any given stimulus there will be a large number of sensory and postural signals. The PPC has a large set of *basis functions* that it applies to these signals. I don’t want to get bogged down discussing the details of these functions: [Zipser and Andersen \(1988\)](#) take them to be linear gain fields, [Pouget](#) takes them to be Gauss-sigmoid functions. For present purposes these details don’t matter (for a good deal of detailed discussion, see [Eliasmith and Anderson 2003](#)). The qualitative idea is that a neuron or neural pool in the PPC will produce a pattern of activity that is some function, a *basis function*, of these sensory and postural signals—for example a Gaussian function of the distance of the location of retinal stimulation from a preferred spot multiplied by a sigmoid function of preferred eye orientation. I will call these entities, these ‘neurons or neural pools,’ PPC-*elements*. Each PPC-element’s activity is the value of some function of the sensory and postural inputs:

$$n_i = B_i(s, q) \quad (1)$$

Here, n_i is the activity of the i th PPC-element, and B_i is the i th basis function, its value on any occasion determined by the sensory s and postural q signals.

There are many different basis functions implemented by the PPC that get computed for each stimulus, each implemented by a dedicated PPC-element, one of the n_i s. While they might all be a kind of Gauss-sigmoid, the difference might be the shape of the Gaussian, the location of its preferred retinal location, and the preferred orientation of the sigmoid. So for a given set of sensory and postural signals, a large number of basis function values of the form specified in (1) will be computed. Each of these basis function values is reflected in the activity of one of the PPC-elements, the n_i s.

What do these PPC-element activities do? They enable certain kinds of motor behavior. In Pouget's model, associated with each of a number of types of basic behavior, such as a *grasp with the left hand*, or a foveating eye movement, is a set of scalar coefficients—think of these as neural connection strengths. So for example a behavior type such as a *left-hand grasp* would have a proprietary and constant set of numbers, g_1, g_2, \dots, g_n , such that when those coefficients are used to produce a linear combination of the PPC-element activities associated with stimulus a , a behavior of that type targeted on stimulus a is correctly executed:²

$$M^{a,g} = (m_1^{a,g}, m_2^{a,g}, \dots, m_p^{a,g}) \quad (2)$$

$$m_j^{a,g} = \sum_{i=1}^n g_{i,j} n_i(a) \quad (3)$$

$$n_i(a)_i = B_i(s_a, q_a) \quad (4)$$

What (2) says is that the neural motor commands that result in a *left-hand grasp* (call this behavior-type g) correctly targeting stimulus a can be represented as a vector $M^{a,g}$, with p components of the form $m_j^{a,g}$. Equation 3 details each of these components. Each is arrived at by multiplying each of the coefficients associated with a *left-hand grasp* (the $g_{i,j}$ coefficients) with the activity of one of the PPC-elements, the n_i s, whose activity corresponds to stimulus a , and adding them together.³ Finally,

² This description is a simplification. There are many 'layers' of processing between the sensory systems and the PPC, and between the PPC and the musculature. As for as the 'correct' motor output, this will be whatever the motor output is that, when sent from the PPC to the 'lower' levels, gets the job done. This does not affect the points I am making here (see Grush 2004b, section R3, for some discussion that, though framed in a different context, is relevant to seeing why this simplification does not affect the main point).

³ It is sometimes common to speak of 'coordinate transformations' in this sort of context, and the sort of mechanisms I am describing here as effecting a coordinate transformation from, for example, retinal space to 'hand centered space'. While not inaccurate, this way of putting things can invite misunderstanding, since it can seem like the output is something like a 'location' relative to the hand or some other effector. What the output is is a location in a 'coordinate frame' for the effector, and this is not a spatial grid centered on the effector, but is rather an abstract space whose coordinates are defined by the effectors' degrees of freedom, in kinematic or dynamic terms. I am simply describing this as a motor command, but no misunderstanding will result in thinking of this as a coordinate transformation so long as what is meant by a 'coordinate' here is kept in mind.

Eq. 4 reiterates that the activity of each PPC-elements resulting from the perception of stimulus a is a basis function of the activities of the sensory and postural signals associated with a .

Of course a left-hand grasp g directed at stimulus b will require a different motor command if b is located at a different spot in egocentric space:

$$M^{b,g} = (m_1^{b,g}, m_2^{b,g}, \dots, m_p^{b,g}) \quad (5)$$

$$m_j^{b,g} = \sum_{i=1}^n g_{i,j} n_i(b) \quad (6)$$

$$n_i(b)_i = B_i(s_b, q_b) \quad (7)$$

Here, the different motor command $M^{b,g}$ —the command that results in a left-hand grasp of stimulus b —is produced by taking *the same set of left-hand grasp coefficients*, the $g_{i,j}$ s, and multiplying them by a different set of PPC-element activations—the ones that the basis functions produce when applied to the sensory and postural signals manifested during the sensing of stimulus b . A different kind of action, like an *eye movement* that foveates stimulus a or b , would be determined in an analogous way: by multiplying the eye movement coefficients ($e_{i,j}$) with the basis functions produced by the stimulus according to the following equations:

$$M^{a,e} = (m_1^{a,e}, m_2^{a,e}, \dots, m_p^{a,e}) \quad (8)$$

$$m_j^{a,e} = \sum_{i=1}^n e_{i,j} n_i(a) \quad (9)$$

$$n_i(a)_i = B_i(s_a, q_a) \quad (10)$$

$$M^{b,e} = (m_1^{b,e}, m_2^{b,e}, \dots, m_p^{b,e}) \quad (11)$$

$$m_j^{b,e} = \sum_{i=1}^n e_{i,j} n_i(b) \quad (12)$$

$$n_i(b)_i = B_i(s_b, q_b) \quad (13)$$

Pouget's model is specific in that it posits a particular kind of basis function, a Gauss-sigmoid function. As such it is a special case of what I will call the *basis function model*. According to the basis function model, the motor commands for behavior types that target stimuli in behavioral space are determined by neural information processing mechanisms that multiply a set of linear coefficients specific to that behavior type with a set of values produced by non-linear basis functions of relevant sensory and postural signals associated with the stimulus. (See [Eliasmith and Anderson \(2003\)](#) for a compatible approach to understanding neural information processing.)

What about the difference between Toni and me? In my case, the sensory signals corresponding to volume and pitch and the rest are not connected to any of the basis function value generation mechanisms in my PPC. All the sounds are there, and neural signals carrying information about the sounds are there, but there is no production of anything corresponding to (1) resulting from them. Rather, on the hypothesis I am

articulating, what experience with the guide does is to allow the PPC to learn how to generate suitable basis function values given the sensory signals that come from the guide and relevant postural signals. When Toni hears Middle C at 35 dB, her PPC automatically combines this signal with various postural signals (especially the orientation of her head with respect to her torso, since the guide is mounted to her head), in order to produce PPC-elements whose activities are capable of being combined with any of the many sets of linear coefficients that are associated with the behavior types she knows. And while these PPC-element activities don't in themselves select for any behavior types, they do specify the details of how any of the behavior types will be executed if they are executed. She is thus in a position to immediately grasp, or orient her head toward, the perceived object, if for whatever reason she chooses to execute one of these types. And the PPC-element activations are the implementation of these detail-specifying dispositions.

This strikes me as a nearly unprecedented convergence of philosophical theory and computational neuroscientific implementation. The Evansian disposition theory clearly identifies the pivot of behavioral spatial purport as the behavioral disposition, and while Evans does not distinguish between type-selecting and detail-specifying dispositions, it seems clear that he had the latter in mind. In the basis function model, we have exactly a model of how sensory and postural signals can be processed in such a way as to yield a set of PPC-element activities that can then be used to produce the detailed execution of any of a large number of basis types. This theory is not only computationally detailed, but has been neurophysiologically vindicated (see [Pouget and Sejnowski 1997](#), [Pouget et al. 2002](#); see also [Eliasmith and Anderson 2003](#)).

3 Amodal emulation, and the emulation theory

3.1 Emulation theory

Here I will very briefly introduce the emulation theory. This introduction will be very schematic and will leave out a great many details and applications (many of which can be found in [Grush \(2004a,b\)](#); the theory is an adaptation of ideas developed in linear control theory, see [Kalman 1960](#), [Kalman and Bucy 1961](#); see also [Bryson and Ho 1969](#) for a standard treatment). I will limit this introductory sketch to three topics: a brief characterization of the basic information processing structure; two paradigm applications; some further remarks on the *dynamic* functions; and a brief discussion of the distinction between modal and amodal emulation. The reader familiar with the emulation theory can safely skip this section, but should not skip Sect. 5.1, where I describe a number of ways in which the theory is misunderstood.

3.1.1 Emulation theory basics

The brains of organisms interact with things, including the organism's body, and its environment. Let's call these things *target processes*, or simply *processes*. A process will change state over time, and typically its state at any time is determined by the following factors: its previous state; its inherent dynamic tendencies; predictable

influences, especially influences induced by the organism itself; and unpredictable influences. If we let a process's state at any time be given by a vector $p(t)$, and simplifying for ease of exposition to linear discrete cases (for generalizations, see Bryson and Ho 1969; to allay concerns about biological plausibility, see Eliasmith and Anderson 2003, who have shown in detail how systems such as those I describe here can be implemented with spiking neurons), then we can summarize the remarks above as:

$$p(t) = Vp(t - 1) + c(t) + d(t) \quad (14)$$

where $p(t)$ is the process's state at time t , $p(t - 1)$ is its state at the previous time $t - 1$, V is a function representing the process's own inherent dynamic tendencies (I will sometimes call V the object's *dynamic*), $c(t)$ is the predictable influence, and $d(t)$ is the unpredictable influence, or *process noise*.

The brains of organisms interact with these processes in two broad ways: there are influences from the brain to the process, and influences from the process to the brain. One direction of influence has already been accounted for: the $c(t)$ is the predictable influence on the process, and the brain's own command signals are known by the brain, and hence 'predictable' in the relevant sense (not unpredictable noise). So for simplicity, from now on $c(t)$ will be my notation for the brain's influence on the process. What about the other direction? Brains have sensors that provide information about the process's goings on, and we can schematically represent this as a *measurement* of the process that results in an observed signal. This observed signal, the influence that goes from the process to the brain, is not perfect, and can be represented conceptually as an ideal measurement to which sensor noise is added:

$$I(t) = Op(t) + n(t) \quad (15)$$

Here, $I(t)$ is the observed signal at time t , O is the measurement function, $p(t)$ is the process's state at time t , and $n(t)$ is sensor noise at time t .

The emulation theory is built around the idea that the brains of organisms construct and maintain internal models, or emulators, of many of the processes with which it interacts, including its body and environment (Ito 1970; Desmurget and Grafton 2000; for more references see Grush 2004a,b). An emulator, when implemented in a brain, is a neural system that the brain can interact with in a way analogous to how it interacts with the process. Consider a toy example based on ship navigation. The process is the ship, particularly the ship's state—its location, heading, speed, and so forth. This process evolves over time as a function of (i) its previous state; (ii) a function, in this case based on physics and fluid dynamics and such, that specifies how this state evolves over time; (iii) predictable influences, most centrally self-generated actions; (iv) unpredictable influences, such as unforeseen winds and currents. The navigation team maintains a model of that process, part of which is the map, but which also includes procedures for updating the map. The result is that this model can be manipulated in a manner analogous to the way that the real process can be manipulated. The captain can issue the command 'right 10 degrees rudder, full speed' to the real process, and thus change the state of ship. But he might also issue a mock version

of that order to the navigation team, who can update the model in such a way that it goes into a state that is a prediction of the state the real process would go into if the command were acted on by the process. (See Grush 2004a,b, for more detail; for interesting application, see Kelly 1994.)

There are several potential uses for such a model that I will mention. First, as described above, the model could be used to try out counterfactuals. Perhaps the captain wants to know if a given command sequence will run the ship aground on a nearby shoal. One source of information is to provide this sequence of commands to the navigation team, but not to the real process. If the navigation team reports that the ship runs aground, then this might be reason to think that if that sequence were actually executed, that would be the result.

A second use is the processing of sensory information. In the case of navigation, the sensory information might be numerical readings provided by 2 or 3 people taking bearings to various landmarks or stars. This sensor signal is subject to noise, meaning in this context that the people taking the bearings are not perfect. One potential way to help minimize the effects of noise is to combine the information from the sensory signal with information provided by the model's prediction. For example, the navigation team plots the state they expect the ship to be in, given its prior state and the current command. And they also plot the state that the sensory measurements say the ship is in. And then they combine the two to provide a better estimate than either source of information alone would produce (see Grush 2004a,b for more information). In principle, the model can help to overcome noise, and even fill in missing sensory information if some of the sensors are intermittent. This is known as *filtering* in the control and signal processing literatures, but in this context it can be thought of as processing sensory information into perceptual information (but see Sect. 5.1).

Very roughly, an a posteriori estimate $\hat{p}(t)$ is arrived at by combining a prediction based upon the previous estimate with the observed signal:

$$\hat{p}(t) = \bar{p}(t) + k\check{p}(t) \tag{16}$$

$$\bar{p}(t) = V\hat{p}(t - 1) + c(t) \tag{17}$$

Here in (16), $\hat{p}(t)$ is the a posteriori estimate, which is arrived at by combining what is expected to happen, also called the a priori estimate, $\bar{p}(t)$, with what was observed to happen, $\check{p}(t)$. Here k is a gain term that describes how the combination is effected. The details of this don't matter for present purposes. As (17) explains, the a priori estimate is arrived at by taking the previous filtered estimate $\hat{p}(t - 1)$, evolving it according to V , which is the knowledge of how the process typically evolves over time, and adding the predictable influence $c(t)$. The *filtered* signal $\hat{I}(t)$, which is an estimate of the observed signal $I(t)$ minus the sensor noise $n(t)$, is arrived at by subjecting the a posteriori estimate to a measurement:

$$\hat{I}(t) = O\hat{p}(t) \tag{18}$$

A third use might be to get quicker feedback than is available from the real sensors. Suppose that the navigation team can plot the next location in 10 or 15 s. They could then also, with a 'measurement' of that state, produce an estimate of the numbers

that the people operating the bearing taking equipment will produce. While this might not be terribly useful in ship navigation, much of the initial motivation, in the 1970s and 1980s, for positing emulators in the human nervous system was as a means of overcoming feedback delays such as this.

Given the distinction between the process and a measurement of that process, there are two broad kinds of ways this can be implemented. Either the emulator emulates the process only, or it emulates the combination of the process and some particular kind of measurement. The first I call *amodal* emulation, the latter *modal* emulation. The navigation example above was an example of amodal emulation. The model modeled the states of the process itself—the ship’s location, heading and so forth. The model could then be subjected to a mock measurement to produce estimates of the observed signal: if the ship’s state were as the model represents it, then the numbers produced by the people taking bearings should be *such and such*, the depth readings should be *so*.

A modal emulator is one that is tied to some modality of measurement. This is not really used in ship navigation, and so the example will be a bit fantastic, but bear with me. One could imagine a navigation team that didn’t maintain a map, but rather learned a lot of relationships between current sensor states, current commands, and subsequent sensor states. For example, they might learn that if the bearing numbers are 121 and 310, and the current commands is right 10 degrees, half speed, then the next set of bearing numbers will be 123 and 301. Granted this would not be the easiest way to do things, but it is an example of how an emulation system might not explicitly represent the process apart from a specific modality. Now of course anyone looking at this system from the outside would know that the reason that this sort of contingency holds has a lot to do with the nature of the ship, physics, the local environment, and so forth. But the team that learns and implements this modal emulator need have no such knowledge. Their ability to emulate the system is tied to a given modality of measurement, and essentially black-boxes everything between the motor output and the sensory signals. Of course, the modal emulator can still be used for all three purposes mentioned above: it can produce predictions (though its predictions will be predictions exclusively about sensory states, not ship locations); it can be used for some kinds of perceptual processing by being combined in one or more ways with the real signal (but see Sect. 5.1); and it could also be used to ameliorate the effects of slow feedback, if necessary.

It will be helpful to flesh out an example of what appears to be an modal emulator employed by the brain, in this case a visual emulator. Duhamel et al. (1992) published findings that seem to point to a modal visual emulator (as suggested by Mel 1986; Grush 1995; Rao 1999). They found neurons in the parietal cortex of the monkey that remap their retinal receptive fields (the area on the retina that a cell is responsive to) in such a way as to anticipate imminent stimulation as a function of a copy of the saccade motor command.

The experimental situation is illustrated in Fig. 1. Box A represents the visual scene centered on a small disk. The receptive field of a given PPC cell is shown in the empty circle in the upper left quadrant. The receptive field is always locked to a given region of the visual space, in this case above and just to the left of the center. Since nothing is in this cell’s receptive field, it is inactive. The arrow is the direction of a planned saccade, which will move the eye so that the new visual scene will be as in B. Before

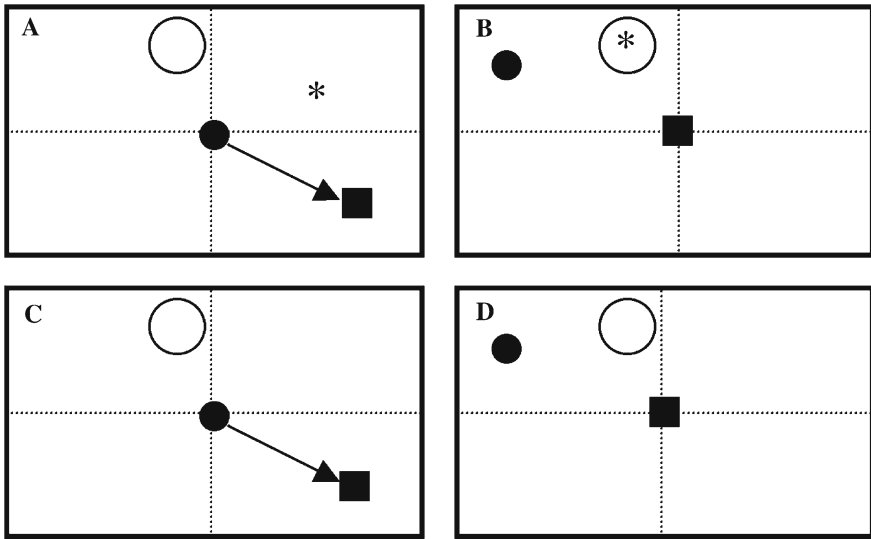


Fig. 1 Anticipation of visual scene changes upon eye movement. See text for details

movement, there is a stimulus, marked by an asterisk, in the upper right hand quadrant. This stimulus is not currently in the receptive field of the PPC neuron in question, but it is located such that if the eye is moved so as to foveate the square, the stimulus *will* move into the cell's receptive field, as illustrated in Box B. The Duhamel et al. finding was that given a visual scene such as represented in Box A, if an eye movement that will result in a scene such as that in Box B is executed, the PPC neuron will begin firing shortly after the motor command to move the eye is issued, but before the eye has actually moved. The PPC neuron appears to be anticipating its future activity as a function of the current retinal projection and the just-issued motor command. That is, it is a visual modal emulator. The control condition is shown in Boxes C and D. In this case the same eye movement to the square will not bring a stimulus into the receptive field of the neuron, and in this case the neuron does not engage in any anticipatory behavior (or, more accurately, it *does* engage in anticipatory behavior, and what it is anticipating, correctly, is that nothing will be in its receptive field, and the appropriate behavior is to not fire). The control condition effectively rules out the hypothesis that the PPC cell is firing merely as a result of the motor command itself. It is only if the motor command will have a certain sensory effect that the PPC cell fires. A full set of these neurons, covering the entire visual field, would constitute a modality specific emulator of the visual scene.

There are two related features of this system that make it a modal emulator. First, there is only one modality that is involved in the emulation, and that is visual. Second, nothing of importance about the process itself—objects in the environment—is being explicitly represented. The entire emulation is confined to what is being sensed in retinotopic areas, and how that sensory stimulation will change as a function of motor commands.

An amodal emulator would be different. It would be an emulator that maintained an estimate of some aspects of the *environment*, say, and then produced new estimates of what will or would be happening in the environment under various conditions (perhaps in part as a result of the organism's actions). So perhaps with an initial estimate to the effect that there is an object *just ahead there*, and a motor command to move the eyes, the result would be that the object is still in the same location. However, when that amodal representation is subjected to a visual measurement, then the result will be different before and after the eyes move. This would be analogous to the navigation team predicting what the person taking the bearing measurement would say before and after he and his alidade are rotated 90 degrees—the landmark is represented as not moving, but the mock measurement can predict that a different bearing number will come in after the rotation. I will discuss amodal emulation in more detail in subsequent sections.

3.1.2 Motor control and motor imagery

The first applications I will discuss concern motor control. Ito (1970, 1984) proposed that the cerebellum contains an emulator (his term was 'forward model,' see Sect. 5.1 for the difference) of the body's dynamics on the grounds that it appeared as though the motor control areas were making adjustments to the control signals on the basis of feedback about the results of the control signals, but before such feedback could actually have been received by the brain from the body. Ito felt that one possibility was that this feedback was being produced by a forward model of the body (plus proprioceptive measurement) such that, when it received a copy of the motor command, the emulator would quickly produce a version of the feedback signal that the body would produce, only the emulator's signal would be much faster in arriving. This debate over the applicability of forward models to human motor control has continued (see [Desmurget and Grafton 2000](#)), and the idea that the brain employs emulators of the body is now a major theoretical position in the physiology of motor control.

Furthermore, a system such as this is well-placed to provide an explanation of motor imagery, the imagined feelings of bodily movement. Exactly the same mechanisms already described are sufficient, provided that the motor command is suppressed from acting on the periphery (see [Wolpert et al. 2001](#); [Kawato 1999](#); see also discussion in [Grush 2004a,b](#)). In such a case the emulator is processing copies of motor commands, and producing the mock proprioceptive signals of the same sort that would be produced during overt action. Both of these issues are discussed in much greater detail in [Grush \(2004a,b\)](#).

3.1.3 Process models

One topic I've not discussed at length anywhere else in my many discussions of emulators is the knowledge represented by the function V in the model. My exposition of the emulation theory typically assumes that there is one process that is being represented, and the function V is the knowledge of how this process will evolve over time. This is a simplification in that often there will be many different things being

emulated, by different emulators, and even within the same emulator, and in such cases the construction of the estimate of what will happen next will involve the application of more than one such function.

What I want to focus on now is the case where the system does not yet know which of two functions (call them V_1 and V_2) is the appropriate one for what it is trying to emulate, because the current observations are consistent with two different processes. Suppose that I sometimes wear glasses and sometimes do not. When wearing them, the motor visual loop is different, in that a motor command to move my eyes a given amount will, when not wearing glasses, foveate an object that is, say 10 degrees left of center, but when I am wearing glasses, the same eye movement will foveate a stimulus that is 12 degrees left of center. And if I wake up from a nap, my visual systems may not at first know which function V_1 or V_2 to employ in order to correctly emulate the visual scene. Before movement, either is consistent with the visual scene (let us suppose). How does the system proceed? Some situations may be such that new data will quickly come in that disambiguates. As soon as I move my eyes, the resulting retinal projection will strongly implicate one of the two models. If both are tentatively running initially, after the first movement one will produce a very high sensory residual (difference between expected and actual observed signal), and the other a very low sensory residual.

In other cases, the disambiguating data might not just be expected to arrive on their own. The system might have to purposefully bring about situations such that the competing functions will be expected to produce very different a priori estimates, hopefully one of them close to, and one far from, the observed signal. Many examples are possible here. To start with a toy example, you might be in an environment with many real, and many fake foam, granite rocks. Supposing they can't be discerned visually, you can easily discern which a given object is by acting on it such a way as to push it into a part of the dynamic range that will produce a high sensory residual for one of the functions. If you physically push something that appears to be a granite rock, the emulator for the foam version using function V_f predicts that the rock will topple over, while the emulator for the granite version, using function V_g , predicts that the object won't budge. You push, the object doesn't budge. V_f produces a high sensory residual, and V_b a very small one. And so the second emulator emerges as the better one to employ in this context. In many cases it might take some degree of skill, gained on the basis of experience with entities that evolve in accordance with different functions, to be able to disambiguate them quickly. Specifically, knowing what parts of the dynamic range of the objects will produce discernible sensory residuals, and also knowing how to get the processes into the relevant regions of their dynamic ranges. This issue will return in Sect. 4.3.

3.2 Basis functions and amodal emulation

Note that the sensory and postural signals that serve as inputs to the basis functions are *observed* signals—unfiltered signals going directly from the process (the body and its sense organs) to the PPC. In the notation of Sect. 3.1, the activities of the PPC-elements, n_i , that determine the behavioral spatial location of the stimulus are determined by basis functions according to:

$$\tilde{n}_i = \tilde{B}_i(\tilde{s}, \tilde{q}) \quad (19)$$

Here I am using the hat notation to indicate an unfiltered signal based purely on observation. Otherwise, (19) is identical to (1). There are two ways in which filtering mechanisms can come into play. First, since the inputs to the basis functions are sensory and postural signals, if filtered (as opposed to unfiltered) sensory and postural signals are processed by these basis functions, the resulting basis function values would be expected to be more accurate:

$$\hat{n}_i = \hat{B}_i(t) \quad (20)$$

$$\hat{B}_i(t) = B_i(\hat{s}(t), \hat{q}(t)) \quad (21)$$

$$\hat{s}(t) = V_s \hat{s}(t-1) + c(t) + k_s \tilde{s}(t) \quad (22)$$

$$\hat{q}(t) = V_q \hat{q}(t-1) + c(t) + k_q \tilde{q}(t) \quad (23)$$

The attentive reader will realize that the materials required for (21), namely the values described in (22) and (23), have already been discussed. The modal visual emulator described in Sect. 3.1.1 precisely is the emulator that produces sensory signal estimates $\tilde{s}(t)$. And the modal musculoskeletal emulator, used for motor control and motor imagery, precisely is an emulator that produces postural signal estimates (the posture is the body's posture) $\tilde{q}(t)$.

I said there were two ways for filtering to play a role in the production of PPC-element activities. The way just described was to get filtered basis function values \hat{n}_i by filtering the inputs to those functions. But there is no reason that the n_i s themselves cannot be emulated, and this is the second way. That is, no reason why the system cannot learn a function V_n that describes how a given set of PPC-element activations, together with a motor command, results in a new set of PPC-element activations:

$$\hat{n}(t) = \tilde{n}(t) + k\tilde{n}(t) \quad (24)$$

$$\tilde{n}(t) = V_n \hat{n}(t-1) + c(t) \quad (25)$$

Until now the activations of the PPC-elements have been described as the result of basis functions applied to sensory and postural signals. But if mechanisms such as those described in (24) and (25) are operative, then in fact these PPC-elements are not tied exclusively to the sensory and postural signals in this way. A set of PPC-element activities *could* be determined by functions of sensory and postural signals, but could also be determined by learning how sets of these values evolve over time (that is, an emulator for how locations in behavioral space change, in part as a function of self-movement $c(t)$), and producing filtered values for the n_i s that way, independently of, or in combination with, the bottom-up process driven by sensory and postural signals.⁴

⁴ And indeed there is another reason to loosen the connection between the activity of a given PPC-element and the value of a basis function of sensory and postural signals—it is plausible to suppose that a stimulus can be perceptually located in the *same* behavioral spatial location through different modalities; that is, reason to think that the same set of PPC-elements could all have the same activations caused by a set of postural signals together with *either* visual sensory signals, *or* auditory signals. I shan't explore this issue further here.

If I am right, the PPC is not limited to the production and maintenance of just one set of PPC-element activities corresponding to its tracking of a single stimulus or object. Visual object tracking studies that investigate the number of moving stimuli that can be simultaneously tracked (Scholl 2001) seem to suggest that people can easily track around 4–6 visual objects. This is presumably indexing the number of distinct sets of n_i s that the PPC can construct and maintain.

4 Disposition theory plus trajectory emulation theory = Skill Theory v2.0, the viable and de-confused successor to the confused Skill Theory v1.x

The basic components are now in place, and I can articulate Skill Theory v2.0. I will do this by beginning with the account of behavioral spatial purport and its basis function implementation as discussed in Sect. 2, and then show exactly what abilities, capacities, and kinds of knowledge come into view as the bare basis-function model is combined with mechanism as discussed in Sect. 3. And for added clarity, I will contrast this at each step with the sort of capacities and abilities that accrue to a system that lacks the initial behavioral spatial purport but has everything else.

4.1 Basis function value computation only

Suppose that there is an environment that is completely empty except for small solid objects that float motionless in the air. And let us imagine two subjects in this environment perceiving it with the sonic guide, Toni and myself. Both sonic guides detect echoes from the objects, and present a complex set of auditory signals through the earphones. Toni's PPC immediately processes these signals as described in Sect. 2. The result is that for each object, there is a corresponding set of activities in a set of PPC-elements, n_i s, in this case basis function values. And in accordance with the disposition theory of behavioral spatial content, the induction of such a set of PPC-element activities constitutes the induction of suitable detail-specifying dispositions, and hence Toni perceives these objects as located in her behavioral space. She is in a position, without cognitive preliminary, to reach out for any of these objects that is within grasp, or point at any that are not, or orient her head directly at it, or execute any of the range of basic actions for which she has a set of appropriate coefficients.

Compare this situation to me and my sonic-guide-induced experience. *Ex-hypothesis* I have not used the device enough to allow my PPC to learn to construct suitable basis function values upon the deliverances of the device. And so I experience sounds without any obvious spatial purport. At best I am hearing a number of distinct sounds, and can discern features of those sounds, like volume and pitch.

Notice, however, that *as so far described* neither Toni nor myself is in a position to predict what the consequences of any movement on our part will be. The mechanisms that determine the activities of the PPC-elements do not, *by themselves*, do anything to yield predictions about how these activities will evolve over time, either on their own or as a result of my own movement. If there is any movement, then a new set of basis function values need to be computed from the new incoming sensory and postural signals.

4.2 Basis functions plus emulation, modal and amodal

Next I want to take into consideration what emulation theory provides, both the modal and amodal versions. As described in Sect. 3, this requires knowledge of one or more functions specific to the temporal evolution of the sensory V_s and bodily V_q (postural) sensor signals, or V_n that describe how the activities of the PPC-elements themselves evolve over time. Such emulators are what provide for the capacity to anticipate the evolution of either modal (sensory) or amodal (PPC-element activities) signals over time, including evolution influenced by behavior.

Let us suppose that I have learned an auditory modal emulator. I might thereby know that if I am currently perceiving Middle C at 35 dB, and if I move like so, the sound will change to High A at 30 dB. I might, with enough of this sort of thing, be able to make very detailed predictions about exactly how the sonic guide's auditory signal will evolve over time as I move around. But this by itself is clearly not to imbue this sound with any spatial significance. And indeed it is not at all clear how *any* amount of this sort of thing could conjure such significance. Yet as we shall see in Sect. 5, some have taken this sort of thing to explain spatial perception. I suspect that the culprit here is mis-judging the lesson from sensory substitution devices. From the fact that the user familiar with a sensory substitution device comes to enjoy spatial content from the deliverances of such a device, and from the fact that among the things familiarity allows the user to do is to learn to anticipate consequences of movement, it is concluded that the latter is sufficient for the former. But as I have tried to show, familiarity also is a matter of learning appropriate basis function value production mechanisms. And this is what is doing the bulk of the spatial lifting here. I will return to this in Sect. 5.

Because Toni's PPC can compute these basis functions that yield PPC-element activations, and is producing a new sets of these activations as she moves around, she is in a position to learn how these sets evolve over time. That is, she has access to the information needed to learn, given a current set of n_i s and a certain sort of movement on her part, what the next set of n_i s will be. This knowledge I have described as V_n . And since these PPC-elements are the vehicles for her representation of the spatial aspects of her perceptual experience, this amounts to her being able to anticipate where something will be in behavioral space as a result of its current location and her current movement.

The difference between Toni, as described above, and myself, as described above, could not be greater. Though we both have learned to emulate, to anticipate the consequences of our own movement, in my case this is restricted to sensorimotor contingencies, and the sonic guide's deliverances remain as devoid of spatial significance as they were before I could predict how they would change. Toni, by contrast, has learned an emulator defined over n_i s, and is thus predicting how the objects' behavioral spatial locations will evolve over time as a function of her own movement.

But I should emphasize that these anticipations of object motion through behavioral space, produced as they are employing only 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 will allow Toni to anticipate the trajectories through behavioral space that result from her own move-

ment, but they 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 .

4.3 Pattern concepts and shape

In the last section I pointed out that a suitable combination of Evans' disposition theory and emulation mechanisms can address the how Toni might anticipate the behavioral spatial consequences of her own movement, but would not by themselves provide knowledge of how objects will move through behavioral space on their own. But clearly we have such knowledge. One aspect of this sort of knowledge concerns how objects can move. A second is about object shapes. These are obviously related since an object's shape can influence how it will move, and how parts can move. The vertex of a cube might, if the cube is spinning, move in very non-rectilinear but yet highly predictable ways. For simplicity I will ignore motion, and relatedly the issue of distinguishing object motion from self-motion. I will focus on shape.

I want to start with a general notion, what I will call a *pattern concept*. This will apply to patterns that are spatial as well as patterns that are not. In the simplest case, pattern concepts can simply be learned by generalizing over experience. For example, if in the toy world I described above the floating objects always came either in groups of eight arranged as vertices of a cube, or in groups of four arranged as vertices of a half-cube tetrahedron, Toni might come to learn this. In what would such knowledge consist? In the expectation that she will not just encounter objects with uncorrelated locations, but in groups of one of the two sorts. In terms of activities of PPC-elements, it will be the expectation that the sets of n_i s that get produced in her perceptual experience come in one of two kinds of groups related in certain ways. There are two features of such knowledge I wish to point out.

First, with such knowledge in hand, Toni might be able to produce a representation of a multi-object shape only some parts of which she can currently sense. For example, her PPC might, upon being presented with sensory and postural signals corresponding to six objects arranged around her in a certain way, produce six sets of n_i s, as basis functions of the sensory and postural signals she is receiving, and then these might induce the cube *pattern concept*, which in turn could induce production of two more sets of PPC-element activities corresponding to the other, unperceived, parts of the cube-group—this is how the knowledge that the n_i s come in certain kinds of sets gets cashed out. In this way, she can come to know that there should be two objects just behind her head that she is not sensing. The n_i s corresponding to these unseen parts would be sufficient to guide a grasping or orienting movement at one of the unperceived but represented objects, and hence she would be representing the unseen vertices as being in specific locations in her behavioral space.

Second, and relatedly, some perceptual situations might be ambiguous. A situation might arise where Toni can perceive four objects arranged around her such that she cannot tell whether she is (i) next to a tetrahedron group-object that she is perceiving in its entirety, or (ii) within a cube group-object that she is perceiving only half of.

The sets of n_i s produced as basis functions of sensory and postural signals could be consistent with both. Toni can use her shape concepts to help her disambiguate. While they are both consistent with what she is *currently* sensing, there are things she is not currently sensing that would indicate which group-object shape she is perceiving. The cube concept, for example, is a set of eight sets of n_i s, four of which are being produced by sensed signals, and four of which are induced by the concept. One of these sets of induced n_i s disposes (in the detail-specifying sense) a certain kind of hand movement. It is indicating that there should be an object just behind her back that she could grasp *like so* (grasp coefficients multiplied by this set of n_i s). She moves her hand as specified and encounters nothing. Situation disambiguated. The object is a tetrahedron.

4.4 Discussion

Exploratory experience—for example Toni’s explorations when first donning the sonic guide—has a number of related but conceptually distinguishable results. First, experience to the effect that, when there is a certain combination of sensory input and postural signals, a given motor action will meet with some sort of success (whatever success is—it could be grasping a stimulus, avoiding it, orienting towards it), provides her PPC with materials to begin learning suitable basis functions that can, once learned well, allow her PPC to produce n_i s that can enable her to guide successful action on the basis of sensory and postural signals. The end point of this kind of learning is that sensory modality being imbued with behavioral-spatial purport.

Second, Toni’s exploratory experience allows her nervous system to learn modal emulators, and unlike the sort of learning discussed in the previous paragraph, this is independent of any goals or success. Just learning that a given sensory situation (perhaps with a kind of behavior) will lead to a certain successor sensory situation is the learning of a modal emulator. This is exemplified in the Duhamel et al. result. The two learning situations are entirely distinct, though related. To put it in control theoretic terms, the sort of learning discussed in the previous paragraph is the learning of an inverse mapping from goals to behaviors that will achieve those goals; the sort discussed in this paragraph is the learning of a forward mapping, from current situations and behaviors to successor situations. (See [Grush 2004a,b](#) for discussion.)

With both types of mechanisms in play, Toni is able to not only perceive things as being located in behavioral space (via basis-function-value implemented dispositions), but is able to emulate this space, to produce anticipations of where a perceived stimulus will be in behavioral space (what the new set of basis function values will be) given the current sensory and postural signals and the candidate motor plan, if any. This sort of thing is captured in Eqs. 20–23.

Furthermore, once she has the capacity to produce basis function values in this way, continued exploration will allow her PPC to learn amodal emulators, that is, to provide anticipations of what the resulting PPC-element activities, the n_i s, will be given the current activities and current movement, if any—without having to go through the intermediary of modal emulation. This is the process described by Eqs. 24 and 25, and which marks the transition to amodal emulation.

Finally, her experience allows her to learn concepts for different kinds of patterns, patterns that recur in her experience. And enough experience with them might give her knowledge of how to disambiguate ambiguous scenes. And this knowledge has two aspects: first, given an ambiguous stimulus pattern, knowledge of what sorts of situations will result in predictions based on one, but not the other, of the competing models producing a high mismatch between expectation and observation; and second, knowledge of efficient ways to bring those disambiguating situations about.

These different components interact in complex ways. When the patterns are patterns associated with sets of PPC-element activities, then they are shape concepts. (If not, as would be the case when I learn patterns in the audible stream from the sonic guide, they are simply modal patterns). Learned modal emulation can allow me to anticipate future sensory input, but this will only be imbued with behavioral spatial import the results are processed by basis function value generating mechanisms. Otherwise, it is nothing more than me learning to anticipate that moving forward turns Middle C at 35 dB into High A at 30 dB, for example.

And the use of emulation together with shape concepts can allow me to intelligently explore. Two different shape concepts compatible with a given set of inputs can be processed, by an amodal emulator, to help me decide on an action such that the emulated result of that action for the two possible kinds of shapes yields clearly discernible results. When I imagine sniffing the rock, the emulator for the foam rock and the granite rock produce very similar anticipations. The emulation thus helps me to know that a close sniff is not a suitable way to disambiguate this situation. However, a *shove* behavior input to emulators using the different concepts yields very different results—the foam rock will tip over, the granite rock will not. I then execute the action recommended by the emulation, and disambiguate the situation.

5 Clarification, comparison, and conclusion

In Sect. 5.1 I will make some clarifications about emulators and the emulation theory. There are other ideas floating around in the literature, such as forward models, sensorimotor contingencies, and predictive learning. None of these is identical to emulation, though they are special cases, in different ways; and none of them is embedded within a larger information processing framework that allows them to do the sorts of things their proponents often want them to do. Then, in Sect. 5.2, I compare the overall account to the most fully articulated competitor, Noe's theory. The comparison will allow me to point out the importance of many of the distinctions that have been made in earlier parts of this paper. In Sect. 5.3, I conclude with some remarks on the theory I have tried to articulate, and ways in which I see it being further developed.

5.1 Clarification of emulation and the emulation theory

I will take this opportunity to make some long-overdue clarifications concerning emulators, the emulation framework, and similar ideas floating around in various literatures. I will first make some clarifications about the notion of the emulator, and then to the relationship between emulators and the emulation theory.

An emulator is an entity that implements a certain kind of input–output mapping, namely the same, or close enough for whatever practical purposes are at hand, input–output mapping as some target system. As such, I am purposefully defining emulators in a way that does not take a stand on *what* precisely is being emulated, or *how* it is emulated, beyond the matching of the input–output function. It is a superordinate category covering many subtypes.

First as to *what* is emulated, there are several options, but I will mention two. It might be just the represented domain itself, or it might be that domain together with some specific form of measurement. I have specified these subtypes as modal and amodal emulators. The signal processing/control theoretic notion of a *forward model* is, however, specific. A forward model is modeling specifically the input–output mapping *of the process or plant*, apart from any modality of measurement. So forward models are one type of amodal emulator. At the other end of the scale, a *sensorimotor contingency* is an input–output mapping that is strictly tied to a particular modality of measurement, and is hence one type of *modal* emulator.

Next, moving on to *how* emulators emulate. Again, there are various options. I will discuss two. First, the emulator might simply be a lookup table of remembered past input–output pairs, perhaps supplemented with some means of interpolation. And when given an input, it looks for the closest stored output, or perhaps interpolates between a few close matches. On the other hand, the emulator might be a dynamic model whose components interact in such a way that that interaction can provide the relevant outputs given the inputs. I have called this latter type of emulator an articulated emulator (Grush 2004a,b). In the limit, the emulator might be articulated exactly correctly, meaning that for each variable, parameter, and dynamic relationship between them that obtains in the represented domain, there is an analogous variable, parameter, and dynamic relationship obtaining between components of the emulator, and it is this parallelism that explains the ability of the emulator to emulate the domain. This particular type of an articulated amodal emulator is essentially the control theoretic notion of a *system identification*.

Yet another notion floating about is what Clark (2006) has called ‘prediction learning.’ While it is not clear that Clark has any particular kind of emulator in mind here (it is simply left unspecified in the discussion), it should be noted that as I use the expression ‘emulator,’ it is not to be assumed that the emulator is a product of *learning* as opposed to programming or innate specification. What is important is the implementation of the input–output mapping. Whether it *learned* this mapping or came to implement it in some other way is a question that the notion of *emulator*, as I use it, takes no stand on. And indeed, it seems that for most of the purposes in this debate, there is no need to beg this question. The only thing that is specified by Clark’s notion of ‘prediction learning’—that the predictive capacity came about through a process of learning—seems to me to be the one thing that should have been left unspecified, at least for present concerns. A skill is still a skill if it innate, after all.

So to summarize, I have tried to define the notion of the ‘emulator’ not merely as another synonym for ‘forward model,’ or any other of these expressions, but as a notion that is usefully general. Mechanisms subserving ‘sensorimotor contingencies,’ forward models, and system identifications are all very different special cases of emulators.

So what is the point of defining the core notion at this level of abstraction? This brings us to the topic of the relationship between the *emulation theory* and emulators. The emulation theory is an information processing *framework* that attempts to describe how a system, *one component of which is an emulator*, can use that emulator for a variety of different ends. These ends might include, as a degenerate case, merely producing a prediction. It might involve combining this prediction with the observed signal to process the sensory signal into a filtered perceptual representation. It might involve a system capable of feeding this prediction to the motor system in a particular way to be used to provide faster feedback. There are a variety of uses, each of which requires mechanisms and detail that go beyond the existence of the emulator (or forward model, or whatever) itself, and the goal of the emulation theory is to explain how these uses can be achieved, quite interestingly as specific modulations of a single framework. What makes the general notion of *emulator* as discussed in the previous paragraph a useful generalization is that all of the subtypes (modal, amodal, articulated, look-up table, etc.) can be used for any of the sorts of purposes supported by the broader emulation theory.

So while it is true that whatever it is that implements (or possibly learns) a sensorimotor contingency, as such, as a sort of modal emulator, note that unless it is embedded in something similar to the rest of the information processing mechanisms that the emulation theory describes, it is nothing more than a free-floating anticipation. I have no doubt that proponents of sensorimotor contingencies want them to have, for example, an influence on perceptual content, but unless additional machinery is provided (and I have done this work for them, see [Grush 2004a,b](#)), then it is nothing more than a mechanism that produces a prediction. (Is it used for anything? For providing mock feedback to motor areas? Imagery? Perceptual filling in? If so, how?)

I make these clarifications because it seems to me that the literature that is inspired by notions floating around this conceptual vicinity is vague, unclear, and underspecified. It also seems to me that the emulation theory as I develop it ([Grush 2004a,b](#)) is suited to making the sorts of distinctions that need to be made (amodal emulator, articulated amodal emulator, and so forth) depending on the application at hand. Furthermore, the framework has been developed with the aim of describing *how* these emulators can contribute to perceptual content, the production of imagery, and so forth, while other notions (such as the ‘prediction learner’ and the ‘sensorimotor contingency’) don’t say anything about this. So to the extent they are trying to be used to explain perceptual content (to take but one possible application), they are crucially underspecified.

5.2 Comparison

It may be helpful to compare Skill Theory v2.0 to previous versions. An initial point of comparison concerns the presumed scope. [Berkeley \(1948\)](#), [Evans \(1985\)](#), and myself (now and in earlier publications) have been concerned exclusively with the *spatial* content of perception. Others, including [Cussins \(1992\)](#) and [Noë \(2006\)](#), have made much wider application. In the interest of space I will limit the comparisons to a few brief comparisons with one author, Alva Noë, since it is his work that is most strongly connected with this view these days. The following paragraphs are taken from [Noë](#)

2006 (though they are essentially similar to the synopsis provided in Noë 2004), and they constitute a recent summary of his view (underlining emphasis has been added):

... perceiving is a way of acting. ... Think of a blind person tap-tapping his or her way around a cluttered space, perceiving that space by touch, not all at once, but through time, by skillful probing and movement. This is, or at least ought to be, our paradigm of what perceiving is. ... perceptual experience acquires content thanks to our possession of bodily skills. *What we perceive* is determined by *what we do* (or what we know how to do); it is determined by what we are *ready to do*. ...

To be a perceiver is to understand, implicitly, the effects of movement on sensory stimulation. ... An object looms larger in the visual field as we approach it, and its profile deforms as we move about it. A sound grows louder as we move nearer to its source. ... As perceivers we are masters of this sort of pattern of sensorimotor dependence. This mastery shows itself in the thoughtless automaticity with which we move our eyes, head and body in taking in what is around us. We spontaneously crane our necks, peer, squint, reach for our glasses, or draw near to get a better look (or better to handle, sniff, lick or listen to what interests us). The central claim of what I call *the enactive approach* is that our ability to perceive not only depends on, but is constituted by, our possession of this sort of sensorimotor knowledge.

[An] implication of the enactive approach is that we ought to reject the idea—widespread in both philosophy and science—that perception is a process *in the brain* whereby the perceptual system constructs an internal representation of the world. No doubt perception depends on what takes place in the brain, and very likely there are internal representations in the brain (e.g. content-bearing internal states). What perception is, however, is not a process in the brain, but a kind of skillful activity on the part of the animal as a whole.

There is much in this that I believe is right—or better, there are several different things going on that are, individually, right. But they are confused with one another in various ways, and the resulting amalgam ends up not only being false, but invites some further false consequences. To make things easier I have underlined several key phrases from these passages. I will discuss each but not in order:

1. To be a perceiver is to understand, implicitly, the effects of movement on sensory stimulation. This is a very simplified version of emulation theory, specifically, it is a description of a modal emulator considered in isolation from the rest of the emulation theory (see Sect. 5.1 above; see Grush 2004a,b). And that is OK. Noë and I are on the same page here. Emulation is necessary for perception. It is not clearly specified how merely having a prediction on board makes one a perceiver since something needs to be said about how this prediction can make a contribution to perceptual content (see 5.1 above), but we can charitably interpret Noë as implicitly embracing the relevant aspects of emulation theory here. Another problem will arise when we come to the question whether perception is representational (more on which below). But the current important point is that, as I have argued earlier in this paper, emulation, while

necessary for perception, is not sufficient for spatial purport. I could learn to anticipate the auditory consequences of my movement when wearing the sonic guide, and yet fail to experience its deliverances as anything but non-spatial sounds. All the sensorimotor contingencies in Heaven and Earth don't add up to a location in behavioral space. They just add up to someone who is very good at, for example, predicting how the sounds produced by the sonic guide will change. What is needed is the disposition theory, and its provision of *amodal* emulation.

2. *What we perceive is determined by ... what we are ready to do.* If this is meant to express *detail-specifying dispositions*, as I've specified with the basis function model and disposition theory in Sects. 2 and 3, then obviously I think it is right. As I have argued above, the spatial content of perception is a matter of the induction of certain kinds of detail-specifying dispositions, implemented as sets of n_i activities of PPC-elements. Note however that this is a process conceptually distinct from emulation. Either could be manifested without the other, though in fact they go together in normal humans. Furthermore, while this will work for bare behavioral spatial purport, it doesn't work for shapes (which require 'spatialized' pattern concepts, as I described in Sect. 4). Whether I perceive the four objects as vertices of a partially observed cube depends on me having the correct pattern concepts, and it is these that, if they are playing a role in a larger system as described in Sect. 4, will induce the dispositions that allow me to perceive them *as* parts of a cube. But to describe this as 'being determined by what I am ready to do' is misleadingly simplified at best.

3. *What we perceive is determined by what we do.* If there is anything right here, it is not the same right thing that is captured in (1) and (2) above, since neither of those need involve any actual action. If it means something like whether I perceive the sky or the mud puddle is determined by what I do (turn my eyes upward or downward), then of course it is right, but trivial. I think that if it means something that is correct, not trivial, and not just a confused way of re-expressing (1) or (2), then it is what I have tried to capture by the disambiguation of ambiguous stimuli (discussed in Sects. 3 and 4). The fact that I perceive the four objects *as* the seen half of a cube, as opposed to a completely seen tetrahedron (see the example in Sect. 4) depends on my having actually reached behind me to check for the disambiguating object.

4. *What we perceive is determined by what we know how to do.* This also strikes me as hopelessly vague. If 'know how' means we have some sort of emulator (as in (1) above), then yes. If it means that we have learned to construct sets of n_i s to guide behavior as in (2), then yes. If it means that we have knowledge of what conditions can disambiguate ambiguous perceptual scenes, and how to bring those conditions about, (as described in Sects. 4.3 and 4.4), then yes. These specifics are all clear, clearly different, and individually correct as descriptions of part of what is involved in perceiving space. If it means anything else, then I lose confidence that it is right.

5. *We ought to reject the idea—widespread in both philosophy and science—that perception is a process in the brain whereby the perceptual system constructs an internal representation of the world.* If anything, this is the opposite of what follows. I have tried to indicate that if Noë's account can be rescued it is by being seen as an instance of Skill Theory v2.0. But neither of the main components of this theory—emulation theory and disposition theory—are inconsistent with a representational theory of perception. Emulation theory, one specific version of which Noë embraces albeit in vague

language, has been touted as a paradigm representational account (see e.g., Grush 1995, 1997, 2004a,b). The emulator is representing the body and/or environment in the same way that a flight simulator represents an aircraft. If we restrict attention to modal emulation (as Noë seems to want to do), it might seem less obvious that it is representational, but then as I have argued above, the result is not spatial perception at all, just well-predicted sensations. And the disposition theory too is representational. Locations in behavioral space are *represented as such* in virtue of the induction of dispositions implemented in the activities of PPC-elements, in the first instance as basis functions of sensory and postural signals.

I have been critical of Noë here, but I could as easily have been as critical about any of the proponents of previous versions of the skill theory, myself included. I dogpile on Noë not because he has been more guilty than anyone else of the vagaries and confusions I have been trying to diagnose and cure, but because he has engaged in the most sustained recent attempt to develop the position. So there is as much compliment as criticism here, though I suppose it's possible that he won't see it that way.

I should also point out that Skill Theory v2.0 is only meant to be a (partial) theory of the *spatial content* of perception. Noë's view is meant to apply to much more than this, perhaps all aspects of perceptual content. It could very well be the case that various elements of Noë's view that I think are confused, or inappropriately applied in the case of spatial purport are perfectly clear and adequately applied when the topic is other aspects of perception or perceptual experience. I take no stand on that here. And it may also be that the best way to read what I am proposing here, if it is on the right track, is as a friendly amendment, a tidying up, of Noë's (and Evans', and Cussins', and my own previous) remarks on spatial import.

5.3 Conclusion

What I have tried to articulate here is an initial sketch, not a complete theory. A complete theory would do many things I have not done. More should be said about shape, about intuitive physics, about the relation between these and other disposition-related aspects of perception or affordances. More should be said about how these basic kinds of content play a role in a system that has the capacity for more complex kinds of content, such as objective, allocentric contents, or egocentric contents with much larger scope than seem to be related to any sort of behavior. More should be said about the relation between distinguishing self-motion from object motion. More should be said about the different reference frames used to guide motor behavior, such as head-centered 'space.' And when the emulation theory is generalized to the trajectory estimation version (see Grush 2005, 2007b), then novel and explanatorily potent elements emerge, especially with respect to the representation of behavioral *time*. Did I mention I'm working on a book?

My current goal has been not completeness, but clarity. Progress on this issue has been hamstrung because, at its foundation, the basic ideas underlying the members of the family of motor theories, skill theories, enactive theories, etc., of spatial import have rested on confused foundations. I believe that the apparatus I have here developed separates out the different fundamental components in a way that not only provides the

best current account of the basic phenomenon—the fact that perception has behavioral spatial purport—but does so in a way that is adequate to the project of sound future expansion.

Acknowledgements I would like to thank Chris Eliasmith, Jon Opie, and Jakob Hohwy, for extremely well considered and useful suggestions and criticisms. The final version of this paper owes a good deal to each of them, even if I was unable, for one reason or another, to follow all of their suggestions.

References

- Aitken, S., & Bower, T. G. R. (1982). The use of the sonicguide in infancy. *Visual Impairment and Blindness*, 76, 91–100.
- Berkeley, G. (1948). An essay towards a new theory of vision. In A. A. Luce, & T. E. Jessop (Eds.), *The works of George Berkeley, Bishop of clyone*, Vol. 1. London: Thomas Nelson and Sons, Ltd.
- Bower, T. G. R. (1977). Blind babies see with their ears. *New Scientist*, 73, 255–257.
- Bryson, A. E. Jr., & Ho, Y. C. (1969). *Applied optimal control: Optimization, estimation, and control*. Waltham, MA: Blaisdell.
- 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, 2594–2606.
- Clark, A. (2006). Cognitive complexity and the sensorimotor frontier. *Proceedings of the Aristotelian Society, Supplemental Volume*, 80(1), 43–65.
- Cussins, A. (1992). Content, embodiment and objectivity: The theory of cognitive trails. *Mind*, 101(404), 651–688.
- Desmurget, M., & Grafton, S. (2000). Forward modeling allows feedback control for fast reaching movements. *Trends in Cognitive Sciences*, 4(11), 423–431.
- Duhamel, J.-R., Colby, C., & Goldberg, M. E. (1992). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science*, 255(5040), 90–92.
- Eliasmith, C., & Anderson, C. (2003). *Neural engineering: Computational, representation, and dynamics in neurobiological systems*. MIT Press.
- Evans, G. (1985). Molyneux's question. In G. Evans (Ed.), *The collected papers of Gareth Evans*. London: Oxford University Press.
- Grush, R. (1995). *Emulation and cognition*. PhD Dissertation, UC San Diego Cognitive Science and Philosophy, UMI.
- Grush, R. (1997). The architecture of representation. *Philosophical Psychology*, 10(1), 5–25.
- Grush, R. (1998). Skill and spatial content. *Electronic Journal of Analytic Philosophy*, 6(6). (<http://www.ejap.louisiana.edu/EJAP/1998/grusharticle98.html>)
- Grush, R. (2000). Self, world and space: the meaning and mechanisms of ego- and allocentric spatial representation. *Brain and Mind*, 1(1), 59–92.
- Grush, R. (2004a). The emulation theory of representation: motor control, imagery, and perception. *Behavioral and Brain Sciences*, 27(3), 377–396.
- Grush, R. (2004b). Author's response: Further explorations of the empirical and theoretical aspects of emulation theory. *Behavioral and Brain Sciences*, 27(3), 425–442.
- 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.
- Grush, R. (2007a). Berkeley and the spatiality of vision. *Journal of the History of Philosophy*, 45(3), 413–442.
- Grush, R. (2007b). Space, time and objects. In J. Bickel (Ed.), *The Oxford handbook of philosophy and neuroscience*. Oxford University Press.
- Heil, J. (1987). The Molyneux question. *Journal for the Theory of Social Behavior*, 17, 227–241.
- Ito, M. (1970). Neurophysiological aspects of the cerebellar motor control system. *International Journal of Neurology*, 7, 162–176.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(d), 35–45.

- Kalman, R., & Bucy, R. S. (1961). New results in linear filtering and prediction theory. *Journal of Basic Engineering*, 83(d), 95–108.
- Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology*, 9, 718–727.
- Kelly, A. (1994). A 3D state space formulation of a navigation Kalman filter for autonomous vehicles. Technical Report, CMU-RI-TR-94-19, Robotics Institute, Carnegie Mellon University.
- Mel, B. W. (1986). A connectionist learning model for 3-d mental rotation, zoom, and pan. In *Proceedings of Eighth Annual Conference of the Cognitive Science Society*, pp. 562–571.
- Mel, B. W. (1988). MURPHY: A robot that learns by doing. In *Neural information processing systems* (pp. 544–553). New York: American Institute of Physics.
- Noë, A. (2004). *Action in perception*. Cambridge MA: MIT Press.
- Noë, A. (2006). Précis of *action in perception*. *Electronic Journal*, 12(1).
- Pouget, A., & Sejnowski, T. (1997). Spatial transformations in the parietal cortex using basis functions. *Journal of Cognitive Neuroscience*, 9(2), 222–237.
- Pouget, A., Deneve, S., & Duhamel, J.-R. (2002). A computational perspective on the neural basis of multisensory spatial representation. *Nature Reviews: Neuroscience*, 3, 741–747.
- Rao, R. P. N. (1999). An optimal estimation approach to visual perception and learning. *Vision Research*, 39, 1963–1989.
- Scholl, B. J. (2001). Objects and attention: the state of the art. *Cognition*, 80, 1–46.
- Wertheimer, M. (1912). Experimentelle Studien über das Sehen von Bewegung. *Zeitschrift für Psychologie*, 61, 161–265.
- Wolpert, D. M., Ghahramani, Z., & Randall Flanagan, J. (2001). Perspectives and problems in motor learning. *Trends in Cognitive Sciences*, 5(11), 487–494.
- 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.