

Perception, Imagery, and the Sensorimotor Loop.

To appear in *A consciousness reader*, Esken and Heckmann (eds) Schoeningh Verlag

Rick Grush

Word Count (including abstract, notes, and references): 7945

### **Abstract:**

I have argued elsewhere that imagery and representation are best explained as the result of operations of neurally implemented emulators of an agent's body and environment.<sup>1</sup> In this article I extend the theory of emulation to address perceptual processing as well. The key notion will be that of an emulator of an agent's *egocentric behavioral space*. This emulator, when run off-line, produces mental imagery, including transformations such as visual image rotations. However, while on-line, it is used to process information from sensory systems, resulting in perception (in this regard, the theory is similar to that proposed by Kosslyn (1994)). This emulator is what provides the **theory** in theory-laden perception. I close by arguing briefly that the spatial character of perception is to be explained as the contribution of the egocentric behavioral space emulator.

### **0. Introduction.**

The intuitive link between perception and imagery is currently being vindicated. There is increasing neuropsychological evidence demonstrating that many of the same cortical areas, including primary sensory areas, are involved in both processes. While this may be more exciting than it is surprising, it is surely surprising to many that most types of imagery also involve increased activity in cortical and cerebellar structures primarily concerned with *motor control*. Much of this article will be aimed at explaining how and why motor areas and sensory areas interact so as to produce imagery. The lessons of this hypothesis stretch far beyond a simple explanation of the neurobiological foundations of imagery to shed significant light on the nature perception, cognition, representation, and ultimately, conscious experience itself. Dealing with all of these would obviously be unrealistic, so I shall limit my ambitions to some remarks on perception.

This article will be organized as follows. Section 1 will briefly introduce an architecture for biological motor control. I will then explain how this motor control architecture, with trivial modifications, can provide for motor imagery. This hypothesis leads to certain predictions about the relationship between motor performance and imagery. Section 2 recounts psychological and neurobiological evidence to the effect that these predictions are borne out. Section 3 generalizes the framework to cover visual imagery, and then quickly runs through some of the evidence that visual imagery is best explained by the same mechanisms. Finally, Section 4 applies this framework to the issue of perceptual processing, and especially the spatial character of perceptual experience.

---

<sup>1</sup> See Grush (1995, 1997)

## 1. Motor Control

In this section I will introduce a theoretical control architecture for motor control. This discussion will be uncomfortably brief.<sup>2</sup> I want to restrict attention to a particular class of movements, fast, voluntary, goal directed movements (excluded are, e.g., reflexes and movements characterized by repeated patterns, such as walking or chewing). This class of motor behaviors is of interest because it presents a dilemma: on the one hand, the movements require feedback (I will limit the discussion to proprioceptive feedback) in order to gain accuracy. On the other hand, because these movements are fast, and because of limits to how quickly signals can be transmitted by neural means, many such motor behaviors may need to sacrifice the relatively slow peripheral feedback.<sup>3</sup> There is, in short, a speed/accuracy trade-off, forced by neural signal transduction limits.

One way to finesse this problem is for the central nervous system to construct and maintain an internal model, or *emulator*, of the body. We can think of the body as performing an input-output mapping from initial dynamical states and motor commands to proprioceptive sensory states. This means no more than that the exact character of the proprioceptive information that the body produces (mostly via mechanoreceptors in the muscles and tendons) depends upon the motor commands sent to the musculature. An emulator of the body would be a neural circuit which, upon receipt of a copy of the motor command sent to the body (an efferent copy), generates a mock version of the proprioceptive signal which the real body will produce while that command is being executed.

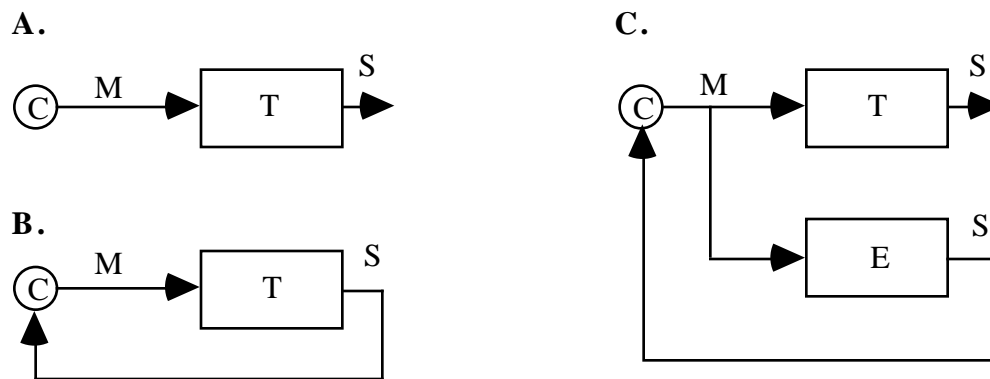
If the emulator's feedback is in the same format as the feedback the body normally produces (proprioceptive, in this case), and if this feedback is available more quickly, then the motor control centers will be able to use the emulator's feedback instead of the real feedback from the body in order to modify its continuing motor commands. If the emulator is realized in neural circuitry which is very near the motor areas, then its feedback might very well be available faster than the real feedback from the periphery. In such a case, the central nervous system would be able to implement pseudo-closed loop control (see Figure 1c).<sup>4</sup>

---

<sup>2</sup> More details can be found in Grush (1995, chapters 2 and 3), (1997), (in preparation).

<sup>3</sup> For discussion of these issues, see van der Meulen et al. (1990), Dennier van der Gon, J.J. (1988), Wolpert et al (1995).

<sup>4</sup> I cannot here go into any of the evidence for the neural implementation of this architecture, but see Grush (1995, 1997). For similar proposals in the literature, see Wolpert et al (1995), Gerdes and Happee (1994).



**Figure 1.** Three control architectures.

Figure 1 shows very simple schematics for three control architectures. In 1a we have open-loop control, in which the controller, C, issues command signals (M, for motor commands) without the benefit of feedback. In a closed-loop scheme (1b), the controller gets feedback (S, for sensor readings) from the controlled system (or as I shall call it, the *target system*, T), which it can use to modify its command sequence. With pseudo-closed-loop control (1c), the controller gets the benefit of feedback (as with closed-loop control) but this feedback (S', for mock, or predicted, sensor readings) does not come from the target system, but rather from an emulator, E, of the target system. Because the emulator is given a copy of the same input as the target system, and because the emulator's input-output function is identical or at least close, the emulator's output will be similar to the sensor output produced by the target system.

It should be clear that in pseudo-closed loop control, we have all the tools necessary for motor imagery. If the motor command is inhibited so that it does not actually go to the body, but yet the efferent copy is still fed to the emulator, then the emulator will provide pseudo-proprioceptive/kinaesthetic information -- *that's just what it does*. And that is exactly what motor imagery is; the internal generation of proprioceptive experience. It should now be equally clear why motor control centers are active during motor imagery -- they are driving the emulator. Pseudo-proprioceptive information is the emulator's output, and just like the body, it provides this output only as a function of motor input.

## 2. Motor Imagery

The hypothesis that motor imagery is generated by the off-line operation of emulators of the motor apparatus, and that these emulators are constructed from monitoring overt performance, has a number of empirical consequences, both psychological and neurophysiological. The implications come from two facts. First, the emulator of the body is driven by the motor centers, just as the actual body is during overt performance. Thus, motor centers should be active during motor imagery. Second, the internal emulator is constructed and tuned through monitoring overt performance; its goal, after all, is to mimic this performance as closely as possible. This entails that motor imagery should share features with overt performance. Moreover, if practicing a motor behavior generally facilitates more skilled performance by improving the efficiency of the executive motor areas, and if during imagery those same centers are receive feedback similar to the feedback they would be getting with overt practice, then imagined practice should also increase motor skills. Amazingly, all of these predictions are borne out.

During many sorts of *overt* motor activity such as playing with your fingers, not only does primary motor cortex show increased metabolic activity, but so do the supplementary motor (SMA) and premotor areas. However, during mental *simulation* of the same movements, primary motor cortex shows no significant increase in activity, *but SMA and premotor areas do show increased activity*.<sup>5</sup> This would suggest that the efferent copy sent to the emulator originates in SMA or premotor cortex,<sup>6</sup> and that during imagery, the normal efferent pathway is inhibited at or before primary motor cortex.

Moreover, the character of mental imagery mirrors the character of overt performance, which is to be expected if the emulators learn their function by monitoring overt performance. For example, Decety and Michel (1989) found that subjects took the same amount of time to write a text fragment and to imagine writing the same fragment. Furthermore, subjects took the same time to write fragments with their left (non-dominant) hands as when imagining writing with their left hands. Evidence of a different sort of isomorphism comes from phantom limb patients. Vilayanur Ramachandran (personal communication) has found that most phantom limb patients fall into one of two groups; those who can voluntarily control their phantom limbs and those who cannot. It

---

<sup>5</sup> See, e.g., Decety, Sjöholm et al. (1990); Roland, Larsen et al. (1980); Fox et al. (1987); Ingvar and Philipsson (1977)

<sup>6</sup> Jeannerod (1994) reaches the same conclusion, though for different reasons (he is interested not in emulation but in motor imagery). He argues that the signals used for imagery originate in premotor cortex or the basal ganglia.

turns out that in almost all cases, those who cannot move their phantoms experienced a period of pre-amputation paralysis, while those who can move their limbs did not. On the present theory, this is to be expected. If the musculoskeletal emulator learns and updates its operation by monitoring the operation of the musculoskeletal system, then when there is a period of paralysis, the mapping learned is that no matter what the motor command is, the proprioceptive result is 'no movement'. When the amputation occurs without a pre-operative paralysis period, there is no information to contradict the operation of the emulator, and hence no reason for it to change (keep in mind the crucial difference between i) proprioceptive information to the effect that there was no movement, and ii) no proprioceptive information about any movements).

Another surprising line of evidence emerges from the sports physiology literature. One can get better at skilled activities by practicing them overtly, of course. No doubt one reason is that during practice, the motor control centers receive information about errors, and they can adapt such that on future attempts at control these errors are lessened. But it is also the case that *imagined* practice has similar benefits -- that is, one can increase skill at overt motor tasks by simply imagining practice.<sup>7</sup> On the present theory, this is quite expected. Imagination is the result of the operation of an emulator, and the emulator will provide feedback similar to that provided by overt performance. If the feedback in the latter case can be used to adapt the controller, so can feedback in the former.

### 3. Visual Imagery

The last two sections dealt with motor control and motor imagery only. But this should not be taken to imply that the strategy of emulation has application only in these domains. This section will show how the same mechanisms can address visual imagery as well. The easiest way to see this is to consider a connectionist model of visual image transformation developed by Bartlett Mel (1986). This model consists of a virtual robot which has two 'eyes' that are trained on a 3 dimensional wire-frame object, such as a cube or pyramid. Images from the two eyes project onto two retinæ, which consist of a grid of connectionist units, each of which serves as a pixel in the grid. The robot can also move with respect to the object it is viewing. It can move closer to, or farther from it, it can move in a circle around it in either direction. When this happens, of course, the pattern of activity on the connectionist unit pixel grid changes continuously.

---

<sup>7</sup> See Feltz and Landers (1983) for review of extensive literature; also see Yue and Cole (1992), and Jeannerod (1994).

The interesting bit is that each of these connectionist units which make up the retinal grid is given as input not only a small segment of the visual scene, but each also has connections to neighboring units, and receives copies of the motor commands sent. Each such unit thus has the opportunity to learn what the next state of activation will be, given the current state and the current motor command. Once the entire grid of units learns to do this, then the grid can act as an emulator. If one primes the grid with an initial state, and then simply gives it a copy of a motor command, it will predict what the next visual image will be *even with the real motor apparatus and eyes disengaged*. A continuing sequence of motor commands, such as circular motion around the object to the left, will result in a rotation of the internally maintained image in the opposite direction. An efferent copy of a 'move forward' motor command will cause the image to enlarge, etc.

Crucially, the ability to do these mental image rotations could never be learned without the image transformations being linked to motor performance. Connectionist units that merely passively received images would be able to learn nothing, because the same initial state could lead to any of a number of subsequent states without rhyme or reason, many of them contradictory. I want, therefore, to flag the point that ***the model can learn the three-dimensional characteristics of objects*** (how such objects look when rotating or moving closer or farther being among these characteristics) ***only through active sensorimotor engagement with them***. This is a crucial point which will resurface in section 4.3.

Though Mel does not analyze it this way, what the retinal grids have learned to do is to emulate image transformations of 3-dimensional objects -- that is, the grid performs double-duty as a sensory input mechanism, and as a visuo-motor emulator. In this case, the target system is the robot's motor apparatus, the surrounding 3-dimensional environment and its population of extended, rigid objects, and the robot's visual-sensory equipment. Given an initial state of this system (specified as a pattern of activity on the retinae), a motor command sent to it will result in a continuous sequence of future states. The visual grid has simply learned this dependency so that it can be taken off-line, exactly as was the case with the musculoskeletal emulator described in the previous sections.

There are many strands of evidence supporting the theory that in biological systems imagery consists of the primary sensory areas' being run via efferent copies in just the way Mel's model suggests. In the model it unsurprisingly turns out that the speed of image rotation is the same as the speed of overt movement around objects. The temporal profile of human visual image rotation is similarly isomorphic to the time it would take to execute the counterpart overt actions. For

example, Droulez and Berthoz (1990) have found that subjects rotate images at speeds similar to the speed at which they make counterpart orienting movements with respect to them.

Furthermore, there is evidence that visual cortex, including area 17 (V1), is active during visual imagery. Kosslyn et al. (1993) found that not only was primary visual cortex active during visual imagery, but that the regions which would be more active during perception of certain kinds of objects were also more active when imagining those objects. That is, when imagining small objects, areas of the occipital lobe which represent the foveal region are selectively active, while during the imagining of larger objects, areas of occipital cortex representing more peripheral visual areas are active as well.

Even more suggestive is a result from Farah et al. (1992). They tested a patient before and after removal of one occipital lobe. Before the occipital lobectomy, the patient was asked to imagine objects of various sizes getting closer until they were as close as possible without overflowing the boundary of the imagined visual field. The finding was that after the removal of one occipital lobe, the distance increased significantly. In other words, the imagined objects were farther away when they filled the entire imagined visual field. This makes sense, because with only one occipital lobe, the visual field is only half as wide.

I want to mention one final point as a segue to the next section, which is that in Mel's model, it is the same grid of units which acts both as the initial processor of real visual input and as the emulator supporting imagery. Mel even shows that once the units have the capacity to emulate the visual transformations, they perform better at processing the normal visual input. That is, *the capacity to emulate not only supports imagery, but also aids normal perception*. This is not surprising, as the retinae, after learning, will be able to anticipate what the next visual image will be, and this will aid in the processing of degraded input. This cooperation of perception and imagery is the topic of the next section.

## **4. Perception**

### **4.1 Imagery and Perceptual Experience**

I want now to introduce one final, but as it happens necessary, refinement to the pseudo-closed loop control scheme. As I have described it, there are two possibilities: the executive areas get their information *either* from the real target system (sensation) *or* from an internal emulator of this

system (imagery). In this case, even though the two sorts of information will necessarily be *in the same format*, they will be quite distinct, and have different sources. A more interesting and useful possibility is that the CNS employs the same internal model for producing imagery as well as for processing sensory information (this was the case in Mel's model). The executive centers would get their information from the emulator in both cases. An analogy should offer some purchase on this idea.

Suppose that I am a general commanding an army in battle. I remain inside a tent at all times, and all information, including my orders, passes in and out of the tent via messengers. Now it won't be possible for me to issue effective commands if all I am aware of is what the most recent messengers tell me, in part because the messengers might not be entirely reliable, in part because each messenger will only convey very partial information about the state of the battlefield, and in part because by the time the messengers get to my tent, the situation on which they are reporting may have changed. So to help keep everything in order, I have a large tabletop in the tent, and on it is a model of the battlefield, with movable pieces for men, machines, tanks, etc. This model will have a number of uses. In the first place, I can use it to predict future states of the battlefield (this is analogous to the case of fast goal directed movement mentioned above). When I issue a command through a messenger to move batallion A to location X, I don't need to wait for a report, many hours later, that batallion A is in location X. I can rather move the piece *representing* batallion A to location X some time after the messenger leaves (how long to wait is something I must learn from experience). The map can also be run 'off-line' to try out thought experiments: if I move these tanks over here, and the enemy responds by moving this artillery over here, then there would be an opening in their defenses over there. This is entirely analogous to the imagery examples mentioned.

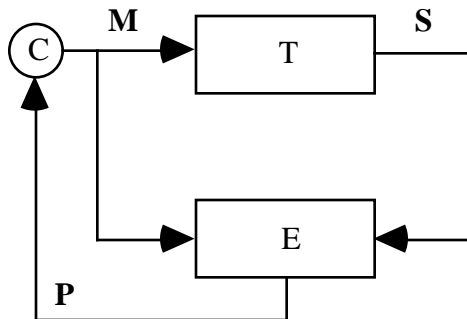
But a final use would be, roughly, the processing of sensory information. That is, given an up-to-date model, reports from messengers can be processed with its aid.<sup>8</sup> To illustrate: if a messenger says that an enemy division has been spotted at location L, I might be able to infer that the enemy has moved his division D (as opposed to division E or F, for instance) from location K to location L, because according to the map/model D is the only division which was close enough to move to L since the last reports. I can update the model accordingly. We might say that although some enemy division in location L was the extent of what was actually *sensed*, the movement of division

---

<sup>8</sup> In this example, and the remainder of this paper, I will speak of a model of the environment being used to aid perceptual processing. But it is not necessary that this model be fully specified -- it can be 'partially elaborated' in the sense of Churchland et al. (1994). That is, the agent might keep much of the environment represented only very schematically. The theory of perception here defended should not be taken as incompatible with the interactive vision approach just because it posits internal representations.

D from location K to location L was *perceived*. Similar 'inferential' mechanisms could help to adjudicate between conflicting reports. The potential for a general account of perceptual filling-in is now within our grasp.

Consider Figure 2, which I will call, for sake of a handy label, a Kalman emulator architecture.<sup>9</sup>



**Figure 2.** Control architecture based on a Kalman filter.

The Kalman emulator has a number of features worth pointing out. Like pseudo-closed loop control (Figure 1c), the controller's feedback comes solely from the emulator. The difference is that the emulator's state is determined not just by the processing of efferent copies, but is influenced by sensor information from the target system as well (if any is available).<sup>10</sup> This means that though the feedback comes exclusively from the emulator during overt performance, that performance is not blind -- what the emulator tells the executive centers is in part dependent on what the sensors tell the emulator. A second less obvious difference is what I am calling the difference between perception and sensation (the 'P' and the 'S' in the figure); a distinction which merits elaboration.

In closed-loop control the target system will be in a certain dynamical state at any given time. But the sensor information used as feedback will typically be sensitive to only a small number of the

<sup>9</sup> The name is thus a bit of a misnomer. The schematic in Figure 2 is based on a control architecture proposed by Gerdes and Happee (1994) for accounting for human movement data. The emulator's function in this case, because it combines sensor information with predicted state information, is based on the Kalman filter (Kalman and Bucy, 1961).

<sup>10</sup> The feed from the target system sensors to the emulator is necessary even in pseudo-closed-loop control when the emulator is being used to predict real feedback (as opposed to when it is engaged in pure imagery). This is because in order for the emulator to predict what the next state of the target system will be, it needs both an efferent copy and *information about the current dynamic state of the target system*.

state variables. These will often be the variables in whose terms the goal state of the target system is specified. But there will be a large number of target system variables which are not sensed. For instance, a car driver is given sensor information about the car's speed, the amount of fuel in the tank, and only a few others, even though there are literally hundreds of other variables of the car which play a role in its operation. The operator's goals are typically expressed in terms of goal values for these few sensed variables (keep the speed between 55 and 60mph, for example). And because a vanilla emulator's only task is to match the input/output function of the target system, its output too must consist of exactly these few variables.

But for the Kalman emulator, matters are different. Its output is not compelled to remain limited to predicted versions of the sensed variables which are the target system's output. This freedom comes courtesy of the fact that the controller in this architecture does not trade-off between the emulator and the target system, coupling with one during imagery and counterfactual reasoning, and with the other during overt performance. Because in Figure 3 the emulator is always coupled to the controller, and the target system is never directly coupled to the controller, the emulator is not forced to narrow its output to the same format as the target system's output. The emulator is free to 'posit' new variables, and supply their values as part of its output. A good adaptive system would posit those variables which helped the controller. We can, without doing violence to the expression, call these variables *theoretical*. They are variables which are not part of the input the emulator gets from the target system. They may be actual parameters of the target system, they may not. But what is important is that *the emulator's output may be much richer than the sensory input it receives from the target system*. For instance, suppose that a car did not have a fuel guage. An entity attempting to predict the operation of the car will obviously do much better if it knows how much fuel the car has, since when this amount reaches zero, the car's dynamic behavior alters significantly. An adaptive emulator of the car might learn to use engine rpm, time, and knowledge of when the last refuelling was to determine a rough estimate of the amount of fuel. But from the emulator's perspective, this would be a theoretical quantity to which it has no direct access.<sup>11</sup>

Plainly, then, this view has the consequence that what one is capable of perceiving can be quite underdetermined by what one is capable of sensing, in a way entirely analogous to how a theory

---

<sup>11</sup> To provide an example in the case of the military commander: if enemy divisions consistently take longer to move from location K to location L than they do to move between other pairs of locations separated by the same distance, then the general might infer that there is a dense forest between K and L, and may update his map accordingly. This would be a theoretical posit, it would become part of the structure of the map, and it would as such play a role in future counterfactual reasoning and perceptual processing. All the same, it is not something that would be directly sensed, and could be wrong.

can be underdetermined by evidence.<sup>12</sup> An immediate lemma is that the deliverances of the sensory apparatus need not accurately mirror what is happening in the environment. As an illustration of this, note that because of the processing abilities made possible by the Kalman emulator, the general can easily make do with scouts who only report *changes of state*, as opposed to constantly reporting on the current state, whether it changes or not. This would be analogous to fast-adapting receptors in biological sensory systems, which are responsive to changes in the feature they are sensitive to, but are relatively insensitive to the absolute value of that feature.<sup>13</sup>

Such considerations about the way in which perception outruns sensation has led Stephen Kosslyn (1994) and Kosslyn and Sussman (1995) argue that imagery is used to enhance perception:

"[We argue] that imagery is used to complete fragmented perceptual inputs, to match shape during object recognition, to prime the perceptual system when one expects to see a specific object, and to prime the perceptual system to encode the results of specific movements." (Kosslyn and Sussman, 1995)

I agree with their ideas, of course, but would not want to describe it as a case of imagery helping out perception, but rather a case of a single mechanism (the Kalman emulator) which has the twin uses of creating imagery and processing sensory information. The difference in description may seem trivial, but I should like it to remain clear that perception and imagery are not two separate processes, such that one may aid the other. Rather, they are essentially the same process, the distinction being the degree, if any, that sensory information is made a significant determinant.<sup>14</sup>

## 4.2 Modalities and the behavioral space.

Kosslyn's program and my own are also separated by another point of emphasis. Kosslyn is concerned with visual imagery, and hence visual perception, almost exclusively, whereas I claim

---

<sup>12</sup> I am here in agreement with Paul Churchland's (1979, 1989) views on the plasticity and theory-ladenness of perception.

<sup>13</sup> For a philosophically interesting discussion of this, see Akins (1996).

<sup>14</sup> A similar idea is to be found in the work of Rodolfo Llinas (Llinas and Pare, 1991). Llinas here argues that cortical activity during dreaming and normal perception is nearly identical, the only difference being that during perception, this cortical activity is modulated by the senses. Llinas even uses the term 'reality emulator'. Since I believe that the phenomenal aspects of dreams are produced by emulators, very little separates Llinas' view and my own.

that emulators support all types of imagery in more or less the same way. I have provided examples of what might be called a proprioceptive emulator, and a visual emulator. And though I hope that by limiting those discussions to single modalities, clarity has been purchased for the introduction of the notion of emulation, I am not satisfied that in the case of perceptual experience having a number of modality-specific emulators provides a viable explanation of the facts.

While it may perhaps be maintained that pure sensory qualia (whatever those might be) have a modality-specific character (sounds are unlike colors), there are clear commonalities between characteristics of visual, kinaesthetic, and auditory *perceptual* experience; and more interesting still there are commonalities between these modes of perceptual experience and *behavioral skill*. When I look at a cup in front of me, I perceive the cup as being a certain distance from me, as being oriented in a certain way, and thus as being graspable in such-and-so a manner. Now if I turn my head (and eyes) to the left a bit, though the character of my immediate visual experience changes, I do not perceive the cup as being in a different location *in my egocentric space*.<sup>15</sup> This is surely because the cup is still graspable by the same motor behaviors, even though my eyes and head may have changed orientation and my visual and vestibular sensory input have changed character. This suggests that perceiving the cup as being at a given egocentric spatial location is at least in large measure constituted by perceiving it as engagable via such-and-such motor behaviors.

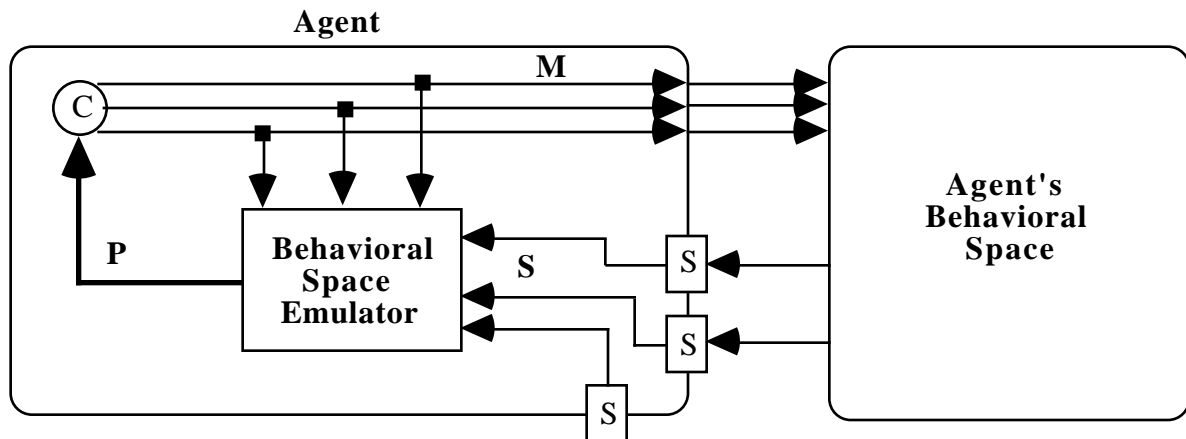
Try engaging in a bit of motor imagery. Without moving, and with your eyes closed, imagine the proprioceptive feeling of your arm stretched straight out in front of you, as though you are pointing at something in the distance. Now what I want you to try to do is to imagine the proprioceptive feel of this experience *apart from the experience of your arm being in a certain spatial location and configuration*. If you are like me, this will not be possible. This proprioceptive experience is part-and-parcel of the egocentric spatial position of your arm, including the *visual* image of what your outstretched arm would look like.<sup>16</sup> Anyone who has ever had their arm 'fall asleep', or otherwise had its proprioceptive feedback interrupted, will attest to how disorienting it is to *see* your arm in

---

<sup>15</sup> Brian Cantwell Smith (1995) exploits the similar insights in an explanation of what constitutes an object. I am in agreement with much of Smith's program, though I want to place emphasis on the notion of a behavioral space whose stabilization perhaps precedes, or is at least co-eval with, the stabilization of objects in that space.

<sup>16</sup> The connection between behavioral space and imagery serves two purposes for me. One is, obviously, an understanding of the processes supporting imagery. The second purpose is to aid recognition of the fact that the behavioral space is internally represented as such. Much work in cognitive neuroscience targets 'coordinate transformations' which would determine what point in one space (visual, perhaps) corresponds to a point in another space (joint-angle space, perhaps). I am in agreement, but am maintaining the stronger position that such transformations are mediated by a single, unifying coordinate system, which is internally represented as the behavioral space, and can thus be run off-line to produce spatial imagery.

such-and-such a position, and yet *not feel* it to be in that position. The point is not just that there are common associations between, e.g., feeling like *this* and looking like *that*. Rather, it is that each of these modalities -- proprio-kineasthetic, visual, and auditory -- together with the behavioral skills involved in negotiating them, gives rise to perceptual experience which has a common spatial element.<sup>17</sup> It is this representation of behavioral space, manifested in the Kalman emulator, which is the stable (though not inflexible) nexus of sensorimotor integration.<sup>18</sup>



**Figure 3.** The Behavioral Space Emulator.

Consider Figure 3. Here we have an agent which acts on its own body (e.g. head and eye movements) and its environment (e.g. manipulating objects) through motor commands sent to

<sup>17</sup> Someone might object, claiming that I have provided no reason why all this cannot be explained by positing associations between characteristics of features from each modality, and hence the inference to a modality-neutral space which is the common ground for all is unwarranted. I cannot adequately address that here. But it is not an embarrassment to my account that the blind, and the deaf, and even those who are both, can perceive spatial relationships and exhibit spatial reasoning not unlike those with their full complement of senses.

<sup>18</sup> Some might be tempted to object that it is not the *representation* of a behavioral space that stabilizes sensorimotor integration, but it is the actual space the agent is in which is the common stabilizing element. But this would be a fairly obvious error, resulting either from insufficient reflection on the facts or from self-deception. The real space the agent is in, and the agent's representation of its egocentric behavioral space, can be doubly dissociated. Dreaming and imagination maintain the integrity of the representation of the behavioral space without reliance on the agent's real location, and neural disruptions of the behavioral space representation, such as hemineglect, disrupt the representation of behavioral space while leaving the agent behaviorally embedded in the same real space it was in before the trauma. The antirepresentationalist's next move, which is to characterize all of this as a context-dependent skill (where one is disrupting either the context or the skill) is no more than an obfuscation of the fact that success in many cases depends on the compatibility of the agent's representation of its egocentric space with the relevant facts about its surroundings. To be sure, there are context dependent skills -- one must be in water in order to swim. One type of context dependent skill is successful representation-mediated activity (depending on context, representations might be inaccurate, and actions based on them unsuccessful). Thus to argue that something is a context dependent skill is not to establish that it is non-representational.

various effectors (the 'M' lines). The agent has a number of sensory organs ('S'), some keyed to its own body (e.g. vestibular sensation), and some keyed to features which are not primarily bodily (e.g. tactile sensation, audition). In addition, the agent has an internal representation of its behavioral space. This emulator is thus in a position to not only predict how the character of sensory experience, *in any of its modalities*, will change as a result of motor commands from any set of relevant effectors, but it is in a position to process sensory information from these various modalities into *perception of a unified spatial environment*. That is, it will be poised to process something which *sounds* like *that* to be something which I could *get a look at* by *moving my head and eyes* like such, and something I might *grasp* by moving my arm in this way.

As a result of this unification, the executive centers are not in the business of trying to manipulate the character of sensory experience, but are rather in the business of trying to manipulate entities in the environment. The leap from the former to the latter (a special case of the leap from being aware of sensory qualia to being aware of things in the environment) is exactly what is afforded by the Kalman emulator of the agent's behavioral space.

### **4.3 *A priori* structure of the behavioral space**

I will now indulge in a slight abuse of terminology. I will use the term '*a priori*' to describe certain features of perceptual experience; features including, but not limited to: spatiality, force-dynamism, and the solidity and temporal continuity of material objects. There are two reasons for taking this use to be a misuse. First, these are features of perceptual experience, so I claim, and are thus not *prior to* perceptual experience. But in this matter I am in agreement with Kant that these features are best thought of as forms of experience, rather than as things experienced -- they thus remain logically prior. The second reason (and at this point Kant will jump ship) is that I think these features are, to some extent at least, constructed by the cognitive system in order to make sense of sensory input. It is thus not necessary that they be either strongly innate or immutable -- two things often associated with '*a priori*'. But I use the term because these features are (how else can it be put?) *supplied to* sensory experience by the cognitive system. Indeed, they are, for the most part, just those features which Hume demonstrated were incapable of being extracted from sensory qualia alone, however hard one squeezes. Such features are a special subset of the theoretical variables posited by the Kalman emulator.

It is undoubtedly fitting that because of limited space remaining for discussion, I must limit the remaining discussion to space. Both of the points I want to make about space I have, in fact, already made. But they are worth explicit attention and another example. The first point is that the

spatial character of perceptual experience is not to be found in sensation (it is thus theoretical in nature), but is a product of the processing which yields perception from sensation. And given the theory of perceptual processing I have posited, this amounts to the claim that the spatial character of perceptual experience is a manifestation of the structure of the behavioral space Kalman emulator. In short, this emulator supplies the spatiality of perceptual experience. And it is this that accounts for the fact that visual space, auditory space, kineasthetic space, and the space in which actions are expressed and evaluated are the *same* space. As Gareth Evans nicely put it, "There is only one egocentric space, because there is only one behavioural space."<sup>19</sup>

The second point, which was broached in section 2, is that active sensorimotor engagement with the environment is a condition for the construction of an emulator capable of coherent spatial processing. The example given was Mel's model of the robot that learned to process and manipulate (in imagination) the spatial properties of objects through active sensorimotor engagement. I will provide another example of this presently, but for now I want to dwell on two implications of this claim. The first implication is that space is a theoretical posit of the nervous system, made in order to render intelligible the multitude of interdependencies between the many motor pathways going out, and the many forms of sensory information coming in. Space is not spoon-fed to the cognizer through the senses, but is an achievement. The second implication is that if the conditions for this multitude of interdependencies are not met, then the cognizer will not develop a representation of space, nor (consequently) spatial perceptual abilities. For instance, if the cognizer is not allowed to actively engage its environment, to explore at least some coherent chunk of the normal range of sensorimotor dependencies, then it will lack any normal understanding of space.

It is with this second implication in mind that I turn now to my final example. It involves not virtual robots, but real biological systems. In a famous experiment, Held and Hein (1963) exposed two kittens to nearly identical visual information. This was done by placing one of the kittens (the passive kitten) in a little gondola, and linking it up to a harness worn by the other (active) kitten so that as the active kitten moved about and explored its environment, the passive kitten was moved in exactly the same manner. The result was that only the active kitten developed normal depth perception. The passive kitten, even though its sensory input was nearly identical, did not .

What I want to claim is that the active kitten could learn an egocentric spatial emulator of its behavioral space, because it had access to both its motor commands as well as the sensory results

---

<sup>19</sup> Evans (1982) *The Varieties of Reference*, p. 160. See also Evans (1985) for a detailed examination of, and argument for, a unified representation of behavioral space.

that those motor commands elicited. This behavioral-space emulator, once appropriately structured, was then used to process sensory information as perceptual information about the kitten's three-dimensional behavioral space. The passive kitten, by contrast, was not able to learn the behavioral-space emulator because of the nigh random manner in which its a sensory field changed -- random because a function of some unknown and unpredictable variable (some other creature's movements).

## 5. Conclusion

Motor activity, imagery, and perception are intimately related, and at the heart of their intersection is representation -- internal emulators that represent the agent's behavioral space. During active exploration of their environments, cognitive creatures construct internal emulators of their environments which are then able to perform the twin functions of processing sensory information (perception), and of allowing disengaged cognition about the behavioral space (including imagery). As Mel's model and Held and Hein's kittens show, even the ability to perceive spatial relations and spatial transformations depends on the construction of a representation of such a behavioral space (though they would not put the point this way, perhaps). It is only within such a behavioral space emulator that there will be a place for notions like *distance*, *persisting object*, *force*, *resistance*, *solidity*, etc.

### Acknowledgements:

I would like to thank the McDonnell Foundation for financial support.

### References:

Akins, Kathleen (1996). Of sensory systems and the 'aboutness' of mental states. *Journal of Philosophy* 93(7):337-72.

Churchland, Paul (1979) *Scientific realism and the plasticity of mind*. Cambridge: Cambridge University Press.

Churchland, Paul (1989). *A neurocomputational perspective*. Cambridge: MIT/Bradford.

Churchland, P. S., Ramachandran, V. S., and Sejnowski, T. J. (1994). A critique of pure vision. In C. Koch and J. L. Davis, *Large-scale neuronal theories of the brain*. Cambridge, MA: MIT Press.

Decety, J. and Michel, F. (1989). Comparative analysis of actual and mental movement times in two graphic tasks. *Brain and Cognition* 11:87-97

- Decety, J., Sjolholm, H., Ryding, E., Stenberg, G., and Ingvar, D. (1990). The Cerebellum participates in Cognitive Activity: Tomographic measurements of regional cerebral blood flow. *Brain Research* 535: 313-317.
- Denier van der Gon, J.J. (1988). Motor control: Aspects of its organization, control signals and properties. in Wallinga et al. eds. *Proceedings of the 7th Congress of the International Electrophysiological Society*. Amsterdam: Elsevier Science Publishers.
- Droulez, J., and Berthoz, A. (1990) The concept of dynamic memory in sensorimotor control. In Humphrey, D.R., and Freund, H.J. (eds) *Freedom to move: Dissolving boundaries in motor control*. Wiley.
- Evans, Gareth (1982) *The Varieties of Reference*. Oxford: Clarendon.
- Evans, Gareth (1985) Molyneux's Question. In *The Collected Papers of Gareth Evans*. Oxford: Clarendon.
- Farah, Martha, Soso, Micheal J., and Dasheiff, Richard M. (1992) Visual angle of the mind's eye before and after unilateral occipital lobectomy. *Journal of Experimental Psychology: Human Perception and Performance* 18(1):241-246.
- Feltz, D.L. and Landers D.M. (1983) The effects of mental practice on motor skill learning and performance: a meta-analysis. *Journal of Sport Psychology* 5:25-57.
- Fox, P.T., Pardo, J.V., Petersen, J.V. and Raichle, M.E. (1987). Supplementary motor and premotor responses to actual and imagined hand movements with positron emission tomography. *Neuroscience Abstracts* 398(10):1433.
- Gerdes, V.G.J., and Happee, R. (1994) The use of an internal representation in fast goal-directed movements: a modeling approach. *Biological cybernetics* 70:513 - 524
- Grush, Rick (1995) *Emulation and Cognition*. PhD Dissertation, University of California, San Diego. At URL <http://www.artsci.wustl.edu/~rgrush/>
- Grush, Rick (1997) The architecture of representation. *Philosophical Psychology* 10(1):5-23.
- Grush, Rick (in preparation) *The Neural Construction of Mind, Language, and Reality. Volume I: Representation, Objectivity, and Content*.
- Held, R. and Hein, A. (1963) Movement-produced stimulation in the development of visually guided behavior. *Journal of Comparative and Physiological Psychology* 56(5):872-876
- Ingvar, D. and Philipsson, L. (1977). Distribution of the cerebral blood flow in the dominant hemisphere during motor ideation and motor performance. *Annals of Neurology* 2:230-237
- Jeannerod, M. (1994) The representing brain - Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences* 17(2):187-202.
- Kalman, R., and Bucy, R.S. (1961) New results in linear filtering and prediction theory. *Journal of Basic Engineering* 83(d):95-108.

Kawato, Mitsuho, Furukawa, K. and Suzuki, R. (1987). A hierarchical neural network model for control and learning of voluntary movement. *Biological Cybernetics* 57:447-454.

Kosslyn, Stephen M. (1994) *Image and Brain*. Cambridge: MIT:Bradford.

Kosslyn, Stephen M., Alpert, Nathaniel M., Thompson, William L., Maljkovic, Vera, Weise, Steven B., Chabris, Christopher F., Hamilton, Sania E., Rauch, Scott L., Buonanno, Ferdinando S. (1993) Visual mental imagery activates topographically organized visual cortex: PET investigations. *Journal of Cognitive Neuroscience* 5(3):263-287.

Kosslyn, Stephen M., and Sussman, Amy L. (1995) Roles of imagery in perception: or, there is no such thing as immaculate perception. In Micheal S. Gazzaniga (ed) *The Cognitive Neurosciences*. Cambridge: MIT/Bradford.

Llinas, R.R. and Pare, D. (1993). On dreaming and wakefulness. *Neuroscience* 44(3):521 - 535.

Mel, Bartlett (1986) A connectionist learning model for 3-d mental rotation, zoom, and pan. In *Proceedings of the Eighth Annual Conference of the Cognitive Science Society*, 562-571. New York: Erlbaum Associates.

Roland, P.E., Larsen, B., Lassen, N.A., and Skinhoj, E. (1980). Supplementary Motor area and other cortical areas in organization of voluntary movements in man. *Journal of Neurophysiology* 43: 1:118-136.

Smith, Brian Cantwell (1995) *The Origin of Objects*. Cambridge: MIT/Bradford.

van der Meulen, J.H.P., Gooskens, R.H.J.M., Dennier van der Gon, J.J., Gielen, C.C.A.M., and Wilhelm, K. (1990) Mechanisms underlying accuracy in fast goal-directed arm movements in man. *Journal of Motor Behavior* 22(1):67-84.

Wolpert, Daniel, Ghahramani, Zoubin, and Jordan, Micheal (1995). An internal model for sensorimotor integration. *Science* 269:1880-1882.

Yue, G. and Cole, K.J. (1992) Strength increases from the motor program. Comparison of training with maximal voluntary and imagined muscle contractions. *Journal of Neurophysiology* 67:1114-1123