An introduction to the main principles of emulation: motor control, imagery, and perception

Rick Grush   Department of Philosophy - 0119
rick@mind.ucsd.edu   UC San Diego
http://mind.ucsd.edu   9500 Gilman Drive
                                              La Jolla, CA 92093-0119

Abstract: A framework for understanding representational capacities of nervous systems is developed and explored. The framework is based upon constructs from control theory and signal processing, most prominently forward models (aka *emulators*) and Kalman filters. The basic idea is that the brain constructs models or emulators for entities with which it interacts, such as the body and environment. During normal sensorimotor behavior these models are run in parallel with the modeled system in order to enhance, supplement and process information from the sensors. These models can also be taken off-line in order to produce imagery, select among possible actions, and solve problems. After introducing the central concepts, the framework is developed in the contexts of motor control, imagery, and perception. Other potential applications, including cognition and language, are briefly explored.

0. Introduction

Throughout most of the 20th century, theories of the neurophysiology of motor control have been dominated by a few simple ideas from control theory: feed-forward (open-loop) and feedback (closed-loop) control. Combinations of these basic ideas have had a reasonable measure of success, but attempts to extend their application beyond motor control to more cognitive domains have failed. The motivation to try to extend them is obvious. Nervous systems evolved to coordinate sensation and action, so to the extent that cognition can be explained as a tweaking or enhancement of mechanisms that subserve sensorimotor function, we will have rendered the phylogenetic emergence of cognitive capacities unmysterious. The basic problem with the assimilation of cognition to sensorimotor behavior has been that cognition clearly involves robust representational capacities, whereas the basic control theoretic tools that have been used to understand sensorimotor behavior do not provide any remotely robust representational capacities.

But a closer look at sensorimotor behavior shows that these basic control theoretic tools are not entirely adequate even to that phenomenon, let alone cognition. Recent work in motor control has suggested that the operation of the nervous system in these domains is better described via more sophisticated constructs from control theory and signal processing. Most prominently for present purposes are forward models (aka *emulators*), pseudo-closed-loop control schemes, and the Kalman filter (e.g. Blakemore et al, 1998; Desmurget and Grafton, 2000; Wolpert et al., 2001; Kawato, 1999) On a feedback control scheme the controller sends motor commands to the body, and gets feedback from various sensors. The controller uses this feedback to modify its motor commands. More sophisticated schemes involve not only a controller, but also a device, an *emulator*, that learns to mimic the input-output operation of the controlled system (the body in this case). These emulators take as input a copy of the motor command, and produce as output a prediction or mock version of the sensory information that the controlled system will produce upon acting on that motor command. These emulators are of great use during sensorimotor behavior, as they can be used to enhance the sensory feedback by filling in missing information, to reduce noise, and to provide feedback that is not subject to feedback delays.

Once in place, emulators can be run entirely off-line by suppressing the real motor command from acting on the periphery, and driving the emulator with an efferent copy. The result of this off-line operation of the emulator is the internal generation of mock sensory

information – in other words, imagery. One immediate use for such imagery is to select a motor plan from a number of candidates by assessing the outcome of each. (Johnson, 2000)

In addition to aiding motor control and providing for imagery, emulators can be used in perceptual processes, especially when part of a *Kalman filter*. A Kalman filter processes sensory information by maintaining an estimate of the states of the perceived system in the form of an emulator. This emulator continually provides a prediction of the future state of the perceived system by simply evolving its most recent estimate. This prediction is then combined with the information from the sensors in order to reach a new estimate based both on the expectation and the sensor information available (see, e.g., Rao and Ballard, 1999).

Like bare feed-forward and feedback control schemes, frameworks employing emulators and Kalman filters have clear application to motor control contexts, and involve only mechanisms that are neurobiologically and evolutionarily unmysterious. But unlike these simpler schemes, they provide for capacities that are genuinely representational, and hence are candidates for supplying the basic representational infrastructure required for cognition.

In section 1 I introduce the basic concepts from control theory and signal processing, focusing on emulators and Kalman filters. I keep to simple control theoretic schemes and discrete linear Kalman filters in order to keep the discussion tractable while providing enough formalism to allow the following discussion to be clear and focused. I close the section with an example (Wolpert et al., 1995) of a model of motor control that uses these mechanisms. In section 2 I turn to motor imagery, and argue that the KF-control scheme introduced in section 1 can explain motor imagery as the off-line operation of an emulator of the musculoskeletal system. I also argue that given what is known about motor imagery, it appears to be the best explanation.

In section 3 I turn to visual imagery. I describe a model of visual imagery (Mel, 1986) that shows how visual imagery can also be explained as the off-line operation of an emulator – in this case an emulator of the motor-visual loop. I then identify two aspects of this explanation of visual imagery; first, that imagery and perception will share at least some processing hardware; and second that covert motor processes will be involved in some forms of imagery. I rehearse evidence that both of these are indeed the case. I also show

how the model can provide for (at least) two kinds of imagery depending on the nature of the emulator that is being run off-line; emulators of the states of modality-specific sensor systems provide imagery that is more experientially vivid, whereas 'higher-level' emulators of the organism's spatial environment and objects in it provide for amodal spatial imagery.

Perception is the topic of section 4. There I show how amodal object/space emulators provide a framework within which merely sensory information can be interpreted. Such emulators not only account for the conceptual distinction between sensation and perception, but, as parts of a Kalman filter, show in detail how processes used in spatial imagery supply the framework within which perceptual interpretation takes place.

In Section 5 I discuss a number of extensions of these ideas, including the relation between amodal imagery as understood on this model, and 'imagery' as used in the so-called imagery debate. I conclude that amodal spatial/object imagery is neither 'imagistic' (narrowly conceived) nor propositional, but has features of both formats. I then briefly point out other potential applications of the framework, including cognition and language. The goal of not to provide detailed mechanisms or knock-down arguments, but to introduce a framework within which a wide range of results and theories in the domains of motor control, imagery, perception, cognition and language can be synthesized in a mutually enlightening way.

1. Motor control: forward models and Kalman filters

1.1 Feed-forward and feedback control

A long-standing controversy in motor control has been the nature of the interaction between the motor centers of the brain and feedback from the body during fast, goal directed movements. (van der Meulen et al., 1990; Desmurget and Grafton, 2000) On one side are those who claim that the movements are ballistic or *feed-forward*, meaning roughly that the motor centers determine and produce the entire motor sequence (sequence of neural impulses to be sent down the spinal cord) on the basis of information about the current and goal body configurations; see Figure 1. This motor sequence is sent to the body which then moves to a configuration near the goal state. It is only at the very end of the movement, when fine adjustments are required, that visual and proprioceptive feedback are used; the bulk of the motor sequence is determined and executed without feedback.
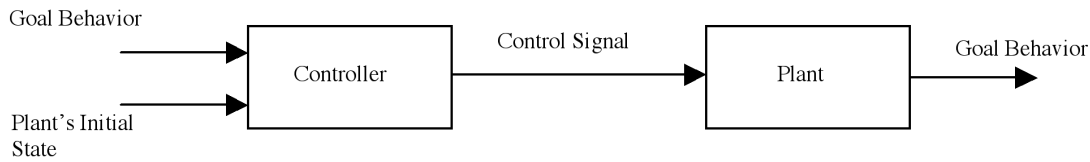
Goal Behavior

Plant's Initial
State

Control Signal

Goal Behavior

Controller

Plant

Figure 1: Feed-forward control. 'Plant' is the control theory term for the controlled system; in the case of motor control, the plant is the body, specifically the musculoskeletal system (MSS) and relevant proprioceptive systems. Here, the 'plant' includes the MSS, sensors, and any random perturbances, such as muscle twitches and noise in the sensors. These components will be treated separately in later sections.

On the other side of the debate has been those who argue for *feedback* control. In the form most opposed to feed-forward control, there is *no* motor plan prior to movement onset. Rather, the motor centers continually compare the goal configuration to the current location (information about which is provided through vision or proprioception) and simply move the current configuration so as to reduce the difference between the current state and the goal until there is no difference between the two. A simplified control schematic for feedback control is shown in Figure 2.
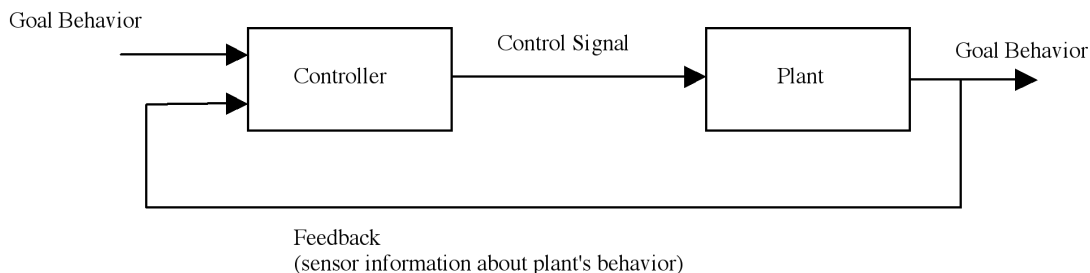
Goal Behavior

Control Signal

Goal Behavior

Controller

Plant

Feedback
(sensor information about plant's behavior)

Figure 2: Feedback control. Sensors (a component of the plant) measure critical parameters of the plant, and this information is continually provided to the controller.

On both schemes, the control process breaks down into two components, the *inverse mapping* and the *forward mapping*. The 'forward' in forward mapping is meant to capture the fact that this part of the process operates in the direction of casual influence. It is a mapping from current states and motor commands to the future states that will result when those motor commands exert their influence. Clearly, this is the mapping implemented by

the actual musculoskeletal system, which moves on the basis of motor commands from its current state to future states. On the other hand, the controller implements the *inverse* mapping. It takes as input a specification of the future (goal) state, and determines as output what motor commands will be required in order to achieve that state. This mapping is just the inverse (or more typically, *an* inverse) of the forward mapping. Hence, when placed in series, a good controller and plant form an *identity mapping*, from goal states to goal states. See Figure 3.
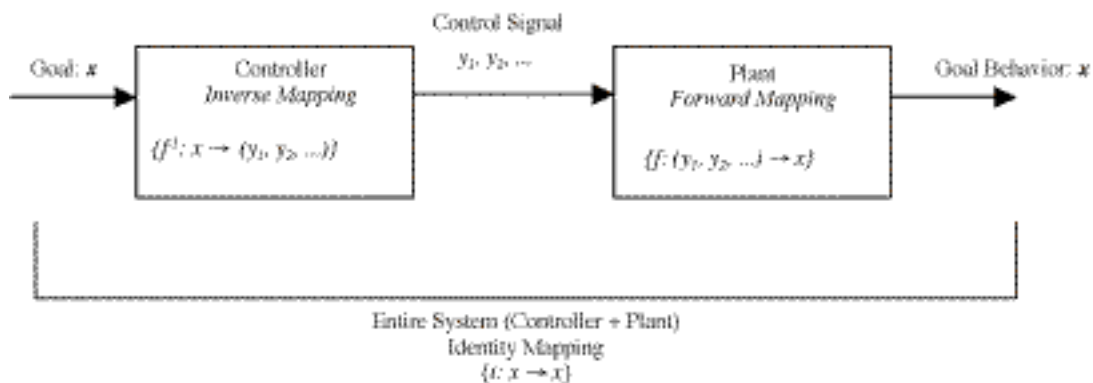

Figure 3. Forward and inverse mappings.

Where the schemes differ is on how the controller produces the control sequence. On the feed-forward scheme, the sequence is produced largely before movement onset. This often requires a good deal of sophistication on the part of the controller. On the feedback scheme, the control sequence emerges over time as the process of interaction between the controller and plant unfolds. But in both cases, the controller produces a motor control sequence, and the body executes it to move into the goal configuration.

Much of the motivation for positing feed-forward control for fast goal directed movements has come from the observation that feedback appears to play little role in movement execution except near the very end of movements (Desmurget and Grafton 2000). Deafferentation (which eliminates proprioceptive feedback), and lack of visual feedback seem not to effect the ability of movements to get most of the way to the target in a roughly accurate manner in much the same way that the initial stages of a movement executed with visual and proprioceptive feedback. The obvious difference shows up at the very end, where lack of feedback severely hinders the ability to tune the end of the movement so as to accurately achieve the target. Tendon vibration studies, in which proprioceptive feedback is distorted, have confirmed this pattern (Redon et al. 1991).

Motivation for positing feedback control has had a number of sources. One has been the seductive theoretical simplicity of feedback control, especially the power and simplicity of servo mechanisms. Another has been findings that appear to show that corrections, made apparently on the basis of information about the state of the movement in the early stages (van der Muelen et al. 1991).

1.2 Emulators (forward models)

But feed-forward and feedback control do not exhaust the alternatives, which is fortunate since a good deal of behavioral data is inexplicable on either model. The data are inconsistent with feed-forward control because the motor sequence is apparently formed, even in early stages, on the basis of feedback. And the fact that corrections are made when peripheral information is not available (van der Muelen et al. 1991) is incompatible with feedback control.

In part for these reasons, there has been growing recognition among researchers in the motor control that schemes involving the use of *forward models* of the MSS are promising (Wolpert et al. 2001; Kawato 1999). I will use 'emulator' as a more descriptive synonym for 'forward model'. In such schemes, the controller exploits a continuing stream of feedback provided by an emulator of MSS dynamics driven by efferent copies of the motor commands sent to the body. The simplest such scheme is shown in Figure 4.
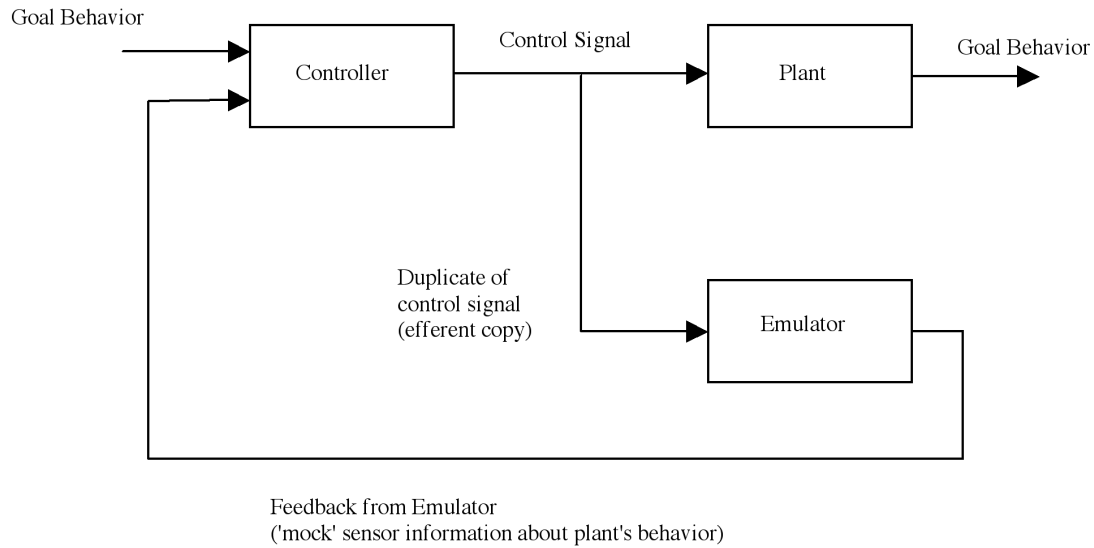
Figure 4. Pseudo-closed-loop control (Ito, 1984). A copy of the control signal, an efferent copy, is sent to a subsystem that emulates the input-output operation of the plant. The emulator thus produces a mock version of the plant's feedback signal.

In this scheme the system includes not just the controller, but also an emulator. The emulator is simply a device that implements the same (or very close) input-output function as the plant. So when it receives a copy of the control signal (it is thus getting the same input as the plant), it produces an output signal, the emulator feedback, identical or similar to the feedback signal produced by the plant. This feedback can be used to guide and correct initial sequences of motor commands until real feedback from vision and proprioception are available at later stages of the movement (Wolpert et al 1995; Desmurget and Grafton 2000) So long as the total loop delay between the controller and the emulator is less than that between the controller and the plant, this scheme will have less of a feedback delay problem, while still producing the right control sequence. Thus the seeming paradox is solved. The reason that there can be corrections to the motor program that are apparently on the basis of feedback, but before peripheral feedback is available, is that the feedback used during these early stages is supplied by the centrally located emulator.

There are two points to highlight that will be important later on. First, for these purposes it does not matter *how* the emulator manages to implement the forward mapping. It might simply be a large associative memory implementing a lookup table whose entries are

previously observed musculoskeletal input-output sequences, and upon receiving a new input, it finds the closest associated output. Another way is for the emulator to be what I will call an *articulated model*. The real musculoskeletal system behaves the way is does because it has a number of state variables (such as elbow angle, arm angular inertia, tension on quadriceps) that interact according to the laws of dynamics and mechanics. Some of these variables are measured, by stretch receptors and Golgi tendon organs. This measurement constitutes bodily proprioception: the 'feedback' in control theoretic terms. Similarly, an articulated emulator is a functional organization of components (*articulants*) such that for each significant variable of the MSS, there is a corresponding articulant, and these articulants' interaction is analogous to the interaction between the variables of the MSS. For example, there would be a group of neurons whose firing frequency corresponds to elbow angle; and this group makes excitatory connections on another group that corresponds to arm angular intertia, such that, just as an increase in elbow angle results in an increase in arm angular intertia, an increase in the firing rate of the first group of neurons instigates an increase in the firing rate of the second. And just as the real MSS is subject to a measurement that provides proprioceptive information, the articulated emulator can have a 'measurement' taken, of the same variables, and thus yield a mock sensory signal.

The second point is that emulators, whether lookup tables or articulated models, must have a certain degree of plasticity. This is because the systems they emulate often alter their input-output function over time. This is *plant drift* – in the case of mechanical plants, belts loosen, gears wear, some parts get replaced by others that aren't exactly the same; in the case of the body, limbs grow, muscles get stronger or weaker over time. Whatever the details, a particular input might lead to one output at one time, but lead to a slightly different output at some time months or years later. In order to remain useful, the overall control system needs to monitor the input-output operation of the plant, and be able to slowly adjust the emulator's operation so as to follow the plant's input-output function as it drifts.

Finally, I should reiterate that the only kinds of motor control with which I am concerned here are fast, goal directed movements. Reflexes and cyclic movements of the sort plausibly produced by the modulation of central pattern generators is not at issue.

1.3 Kalman filters

The advantage of pseudo-closed-loop control is that it is conceptually simple, making it an easy way to introduce certain ideas. It is too simple to be of much use in explaining real biological systems. But the main ideas carry over to more sophisticated and useful schemes.

The next conceptual ingredient to be introduced is the Kalman filter (which I will abbreviate KF; see Kalman 1960; Kalman and Bucy 1961; Gelb 1974). My discussion of KFs here will make a number of omissions and simplifications. For example, I discuss only state estimation/updating and not variance estimation/updating; I discuss only discrete linear models, and ignore generalizations to continuous and nonlinear systems. My goal is simply introduce those aspects of KFs that are important for the remaining discussion. For a more complete discussion, see Grush (forthcoming), and the references therein. The technique, a standard version of it anyway, is diagrammed in Figure 5.
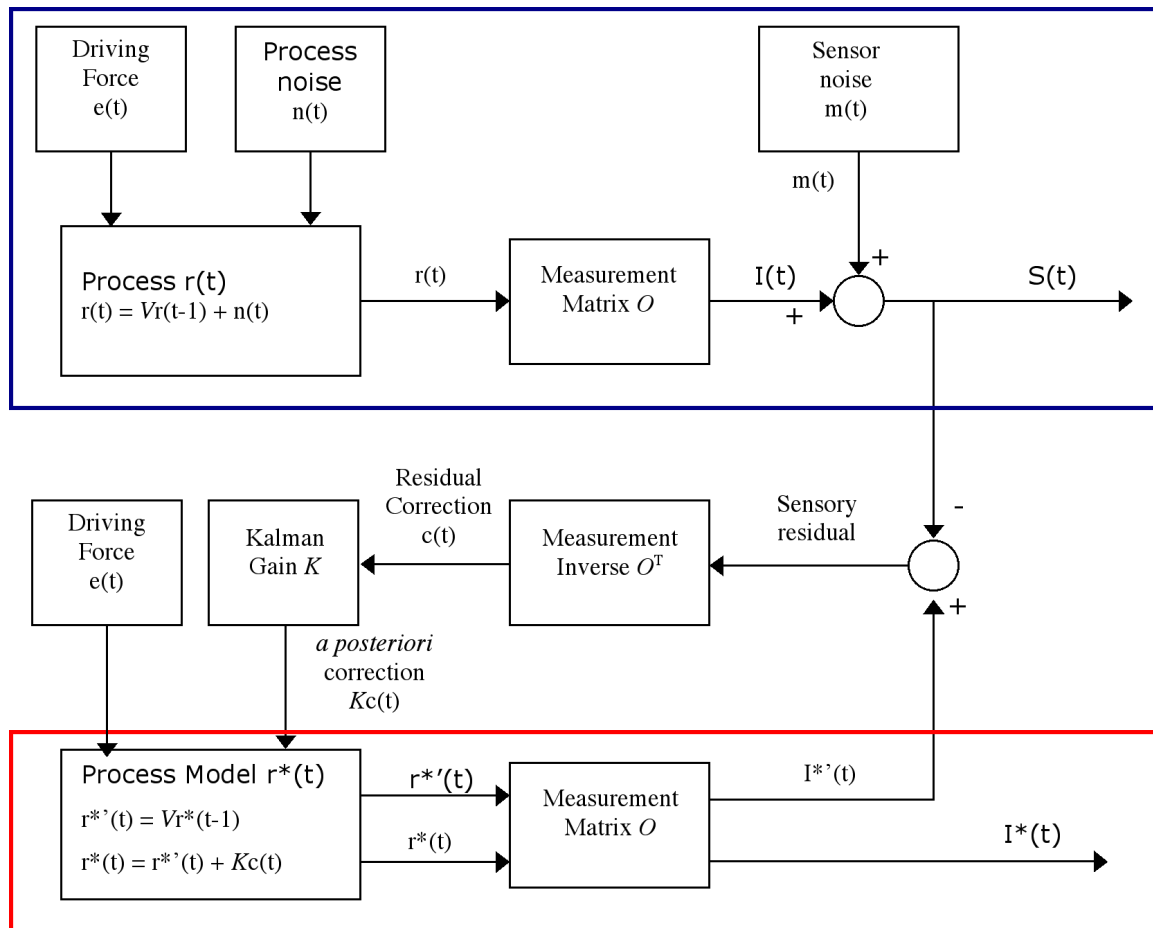
Figure 5. A basic Kalman filtering scheme.

First we need a description of the problem to be solved by the Kalman filter. The problem is represented by the top part of the diagram, within the blue outline. We start with a *process* consisting of a system of k state variables, and whose state at a time *t* can thus be described by the k x 1 vector r(t). The process's state evolves over time under three influences: first, the process's own dynamic (represented here by the matrix *V*); second, process noise, which is any unpredictable external influence; and third, the driving force, which is any predictable external influence. Without the noise of the driving force, the process's state at any given time would be a function of its state at the previous time: r(t) = *V*r(t-1), where *V* is a k x k matrix that maps values of r into new values of r. Noisiness (random perturbations) in the evolution of the process can be represented by a small, zero-mean, time-dependent k x 1 vector n(t); and the driving force as another k x 1 vector e(t). Thus the process at time t is: r(t) = *V*r(t-1) + n(t) + e(t).

The signal *I*(t) is a *measurement* of states of this process. We can represent this measurement as a k x h *measurement matrix O*, that maps r(t) into the h x 1 signal vector I(t): I(t) = *O*r(t). (An obvious special case is where *O* is the identity matrix *I*, in which case I(t) = r(t). This possibility will come up later on.) We can represent unpredictable noise in the measurement process – the sensor noise -- as another small zero-mean h x 1 vector m(t), and so the actual output, the observed signal, is S(t) = I(t) + m(t). The problem to be solved is to filter the sensor noise m(t) from the observed signal S(t) in order to determine what the real, noise-free, signal I(t) is.

Qualitatively, we have a system (the process) whose state evolves over time in such a way that its later states are largely dependent on its earlier states and predictable external influences, but which is also subject to some unpredictable perturbation. For example, the position and momentum of pool balls on a table at time t is a function of their position and momentum at time t-1, predictable external influences such as impacts from cue sticks, and the laws of dynamics that describe how these states evolve over time. The exact position and momentum of the balls is also subject to unpredictable process noise, such as air currents around the table. A measurement of this system – which can be thought of as the product of sensors sensitive to some of that system's states – is produced, but it too is subject to some random noise. The sensors might be low-resolution video cameras trained on the pool table surface, and the signal produced a video image. We want to be able to

filter the sensor noise from the signal in order to determine what the true state of the process system is as time progresses. In the example, determine, on the basis of the low quality video image, the position and momentum of the pool balls.

The core idea behind the KF is that it maintains an optimal estimate of the real process's state, and then measures this state estimate to get an optimal estimate of the noise-free signal. The estimate of the process's state is embodied in the state of a *process model*, an articulated model of the process, also consisting of k parameters. We can represent the process model's state as r*(t). The KF keeps r*(t) as close as it can to r(t), meaning that it tries to match, as closely as possible, each of the k state variables of r* to the corresponding state variables of r. This is done in two steps.

The first step is often called the *time update*. Given the previous state estimate r*(t-1), the KF produces an expectation or prediction of what the state at t will be by evolving r*(t-1) according to the same dynamic *V* that governs the evolution of the real process and adding the driving force e(t).[1] This is the *a priori* estimate, r'*(t) (note the prime), and as stated, it is arrived at thus: r'*(t) = *V*r*(t-1) + e(t). It is called the *a priori* estimate because it is the estimate arrived at before taking information in the observed signal into account. Qualitatively, the KF says "Given my best estimate for the previous state, and given how these states change over time, what should I expect the next state to be?" On the pool table analogy, this would involve letting the modeled pool table balls to roll and collide to see what their positions and momenta become.

The next step, often called the *measurement update*, uses information from the observed signal to apply a correction to the *a priori* estimate. This is done in the following way (roughly – again I must note that my description here is making a number of short cuts). The *a priori* estimate r*'(t) is measured to produce an *a priori* signal estimate I*'(t). E.g., the model pool table is 'measured' to get a prediction of what the image of the real pool table image should look like. This is compared to the observed signal S(t). The difference is called the *sensory residual*. From this residual it is determined how much the *a priori* estimate would have to be changed in order to eliminate the residual altogether; how much would the pool balls in the model have to be moved in order for the video image of the

---

[1] It might be wondered what justification there is for assuming that the driving force can be predicted accurately. This is just by definition. It is assumed that the process is subject to external influences. Any influence that is completely predictable is a driving force, the rest of the external influence – whatever is not predictable – is process noise. So in a case where there were an 'unpredictable' driving force, this would actually be part of the process noise.

model to exactly match the video image from the real table? This is done by pushing the residual through the inverse of the measurement matrix O, O^T. This is the *residual correction.* Though the KF determines how much it would have to alter the a priori estimate in order to eliminate the residual, it does not apply the entire residual correction. Why? The residual correction is how much the a priori estimate would have to be altered in order to eliminate the sensory residual. But though the sensory residual is a measure of the difference between the a priori signal estimate I*'(t) and the observed signal S(t), it should not be assumes that this difference is the result of the a priori estimate's inaccuracy. The a priori estimates r*'(t) and I*'(t) might be very accurate, and the sensory residual due mostly to the sensor noise.

Rather, the KF determines how much of this correction to actually apply, based on the KF's estimates of the relative reliability of the *a priori* estimate versus the noisy observed signal S(t) (the determination of the relative reliability is part of the process I have not gone into, it is the determination of the Kalman gain). To the extent that the process noise is small compared to the sensor noise, the *a priori* estimate will be more reliable, and so a smaller portion of the correction is applied to the *a priori* estimate. To the extent that the sensor noise is small compared to the process noise, the observed signal is more reliable than the *a priori* estimate, and so a greater portion of the residual correction is applied.

Qualitatively, the KF compares its expectation of what the signal should be to what it actually is, and on the basis of the mismatch adjusts its estimate of what state the real process is in. In some conditions, such as when the sensors are expected to be unreliable, the expectation is given more weight than the signal. In other conditions, such as when the process is less predictable but sensor information is good, the expectation is given less weight than the signal.

The result of the measurement update is the *a posteriori* estimate r*(t), which is a function both of the expectation and the observation. This estimate is measured using O in order to get the final estimate I*(t) of the noise-free signal I(t).

1.4 Kalman filtering and control

While KFs are not essentially connected to control contexts, they can be easily incorporated into control systems. Figure 6 shows a Kalman filter incorporated into a control scheme very similar to the pseudo-closed-loop scheme of Figure 4. In Figure 6, everything within the dotted-line box is just the KF as described in the previous section, and shown in Figure 5. The only difference is that the external driving force just is the command signal. Everything in the blue box is functionally equivalent to the *plant* in the pseudo-closed-loop scheme. The box labeled 'plant' in Figures 1 - 4 did not separate out the system, sensors, and noise, but lumped them all together. The 'emulator' box similarly lumped together the emulation of the system and the emulation of the sensors.
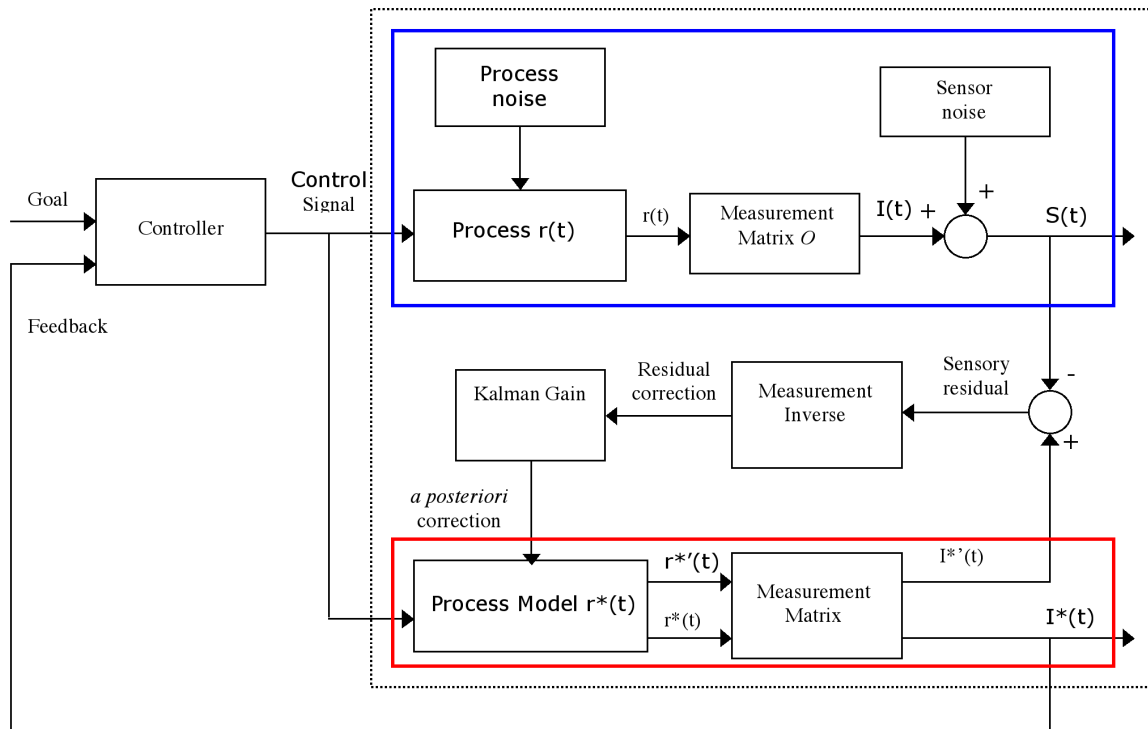
Figure 6. A control scheme blending pseudo-closed-loop control and a Kalman filter.

In effect, this is a control scheme in which an articulated emulator is used as part of a Kalman filter for the purpose of filtering noise from the plant's feedback signal.

Note that the scheme in Figure 6 subsumes closed-loop control and pseudo-closed-loop control as special cases. If the Kalman gain is set to 1, so that the entire sensory residual is

applied, the scheme becomes functionally equivalent to closed-loop control. When the entire sensory residual is applied, the a posteriori estimate becomes whatever it takes to ensure that I*(t) exactly matches S(t). Thus, the signal sent back to the controller will always be whatever signal actually is observed from the process/plant, just as in closed-loop control.

On the other hand, if the Kalman gain is set to 0 so that none of the residual correction is applied, then the a priori estimate is never adjusted on the basis of feedback from the process/plant. It evolves exclusively under the influence of its own inner dynamic and the controller's efferent copies, just as is pseudo-closed-loop control.

I will refer to systems like that in Figure 6 as KF-control schemes, though I will indulge in a certain degree of flexibility in that I will take it that extensions to continuous and nonlinear systems are included, and that the operation may not always be optimal, as when the dictates of the Kalman gain are overridden in order to produce imagery (see Section 2).

1.5 Motor control

The introduction of the basic ideas is complete. The rest of this article will explore applications of these materials to various aspects of behavioral and cognitive neuroscience. The first application is within the domain of motor control, and I will simply report on a model by Wolpert, Ghahramani and Jordan (Wolpert et al., 1995). These authors collected behavioral data from subjects concerning the accuracy of their estimates of the positions of their hands for movements of various durations under three conditions: assistive force, resistive force, and null. The results showed that subjects consistently overestimate the distance that their hand traveled, with this overestimation increasing until a peak is reached at about 1 second. After this time, the overestimate drops to a plateau.

The researchers then develop a model of the sensorimotor integration process, and show that this model closely replicates the time-dependent pattern of estimate and variance errors in all conditions. The model is a KF control scheme essentially identical to the one described above in section 1.4. Because proprioceptive feedback is not available during the initial stages of the movement, the state estimate is based almost entirely on the uncorrected predictions of the forward model. Initial error in this estimate only compounds as time progresses. However as feedback comes from the periphery, this information is used to make corrections to the state estimate, and thus the error drops.

This model is just one of many proposed in recent years that use an emulator (forward model) as part of a system that combines sensory information with an estimate produced by a forward model (e.g. Blakemore et al. 1998; Kawato 1999; Wolpert et al. 2001; Krakauer et al., 1999).

2. Motor Imagery

2.1 The emulation theory of motor imagery

The Kalman gain determines the extent to which the sensory residual influences the *a priori* estimate produced by the emulator – qualitatively, the degree to which raw sensory input trumps or does not trump expectation. Typically the Kalman gain is determined on the basis of the relative *variance* of the prediction and the sensor signal. This is part of what allows the KF to be an optimal state estimator, and in control contexts having optimal information is typically good. The Kalman gain allows us to breathe some much needed flexibility and content into the stale and overly metaphorical distinction between top-down and bottom-up perceptual processing. In the terminology developed in the previous section, we can see that a KF processor is top-down to the extent that the Kalman gain is low -- the lower the Kalman gain, the more the representation is determined by the expectation, which in turn is a function of the operation of the model's inner dynamic as driven by efferent copies, if any. The higher the Kalman gain, the more this inner dynamic is overridden by sensory deliverances. That is, the same system is flexibly top-down and bottom-up, as conditions and context dictate.

Section 4 will explore this in the context of perception. For now, I want to draw attention to the fact that a system that is set up to allow for flexibility in this regard can be exploited for other purposes. Specifically, it can be used to produce *imagery*. Two things are required. First, the Kalman gain must be set to null so that real sensory information has *no* effect. The emulator's state is allowed to evolve according to its own dynamic, as driven by the efferent copies when appropriate; there is no 'correction' or alteration from the senses. Second, the motor command must be suppressed form operating on the body. Thus on this view, motor imagery results from the control centers of the brain driving an emulator of the body, with the normal efferent flow disengaged from the periphery, and with the normal sensory inflow having no effect on the emulator's state and feedback. In this section I will briefly defend this *emulation theory of motor imagery*.

A number of researchers (Johnson 2000; Jeanerrod 2001) are currently favoring a 'simulation' theory of motor imagery, according to which motor imagery just is the inner simulation of overt movement. This is typically cashed out as the operation of motor areas as they would be engaged during overt movement, only with the suppression of the motor command from acting on the periphery. From the point of view of the emulation theory described above, the simulation theory is *half* correct. The part that is correct is that those areas corresponding to the controller -- motor areas -- should be active during motor imagery. Accordingly, the evidence brought forward in favor of the 'simulation' theory is evidence for at least half of the emulation theory. The difference is that the simulation theory does not posit anything corresponding to an emulator; as far as I can tell, the simulation theory is conceived against the backdrop of closed-loop control, and imagery hypothesized to be the free-spinning of the controller when disengaged from the plant. The next section quickly recaps two areas of evidence typically cited in support of the simulation theory. This is evidence to the effect that motor imagery involves the off-line operation of motor control centers. As such, it is also evidence in favor of the emulation theory. In section 2.3, I will discuss considerations favoring the emulation theory over the simulation theory.

2.2 The *simulation* theory of motor imagery

There are two related sorts of evidence cited by proponents of the simulation theory – the first is that many motor areas of the brain during motor imagery, the second concerns a number of isomorphisms between overt and imagined movements.

That motor imagery involves the off-line activity of many motor areas is a widely replicated result (for a recent review, see Jeanerrod 2000). PET studies of motor imagery consistently show selectively increased activity in premotor areas, supplementary motor areas, the cerebellum, among others. This is the defining feature of the simulation theory: that motor imagery is the psychological counterpart of the off-line operation of the brain regions that normally drive motor behavior. Of the major motor areas canvassed in such studies, only primary motor cortex is conspicuously silent during motor imagery. This would seem to imply that the major signal bifurcation where the efferent copy is split from the 'real' efferent signal occurs just before primary motor cortex.

Furthermore, a number of parallels between motor imagery and overt motor behavior have suggested to researchers that the two phenomena have overlapping neural and psychological bases (Jeannerod and Frak, 1999; Deiber et al. 1998). To take a few examples, there are close relations between the time it takes subjects to actually perform an overt action, and the time taken to imagine it (Jeanerrod 1995). The frequency at which subjects are no longer to overtly move their finger between two targets is the same for overt and imagined movement. And there is evidence that even Fitts' Law is in effect in the domain of imagination (Decety and Jeannnerod 1996).

Johnson (2000) has provided compelling evidence that subject's expectations about what range of grip orientations will be comfortable is very close to the range that actually is comfortable. Johnson interprets this result to indicate not only that motor imagery is used for this task (it is motor rather than visual because the crucial factor is the biomechanics of the arm, not its visual presentation), and argues that such imagery, exactly because it respects biomechanical constraints, is used to determine an effective motor plan before execution.

Compatible with Johnson's hypothesis, Hubbard et al. (2000) found that subjects with cerebellar damage are not subject to the size/weight illusion, which causes normal subjects to perceive large objects as lighter than small objects that are in fact of equal weight. Their explanation was that subjects expect larger objects to be heavier, and that normally this expectation is played out in terms of covertly simulated actions of hefting the objects involved. When the result of the covert simulation is compared to the actual felt weight, the large object's felt weight is overestimated compared to the small object's, and hence it feels lighter 'than expected', while the smaller feels heavier. The authors conclude that "the cerebellum may be involved in purely perceptual and cognitive predictions functioning as a *Grush emulator* or forward model for internal simulations before performing certain tasks" (the terminology is due to Hubbard *et al.* 2000).

Thus the proponent of the simulation theory points out that not only are motor areas active during motor imagery, but that the isomophisms between the observed activity of motor areas and the motor imagery suggest that motor imagery is in fact the product of the operation of motor centers, whose operational parameters are tuned to overt performance and hence recapitulated in covert performance.

2.3 Emulation vs. simulation

The evidence I have so far marshaled in this section has been in favor of the hypothesis that motor imagery necessarily involves the off-line operation of motor areas of the brain, those that would be included as part of the 'controller' in a control theoretic block diagram of the motor system. As such, this evidence does not distinguish the simulation from the emulation theories, as both expect the controller to be active during the production of imagery. The difference is that the emulation theory claims that mere operation of the motor centers is not enough, that to produce imagery they must be driving an emulator of the body (the MSS and relevant sensors).

There are reasons for preferring the emulation theory. A bare motor plan is either a dynamic plan (a temporal sequence of motor commands or muscle tensions), or a kinematic plan (a plan for limb movement specified in terms of joint angles). By contrast, motor imagery is a sequence of *faux* proprioception. The only way to get from the former to the latter is to run the motor plan through something that maps motor plans to proprioception, and the two candidates here are a) the body (which yields real proprioception), and b) a body emulator (yielding faux proprioception).

That a motor plan and a sequence of proprioceptive feelings are distinct should be obvious enough, but the difference can be brought out rather nicely by a particular phantom limb phenomenon. Phantom limb patients fall into two groups: those who can willfully move their phantoms, and those who cannot. While not an entirely hard and fast rule, quite commonly those who cannot move their phantoms suffered a period of pre-amputation paralysis in the traumatized limb, while those who can move their phantoms did not, the trauma resulted in the immediate amputation of the limb with no period of pre-amputation paralysis (Ramachandran, personal communication). What follows is a possible explanation for this fact, but I should note that the point I want to draw does not depend on whether or not this possible explanation is correct. The possible explanation is this. Recall the point made in section 1.2 about the requirement that emulators be able to alter their input-output mapping in order to track plant drift. In the case of subjects with a paralyzed limb, the emulator has a long period where it is being told that the correct input-output mapping is a degenerate many-to-one mapping that produces as output 'no proprioceptive change'

regardless of the motor command sent. Eventually, the emulator learns this mapping, and the emulated limb becomes 'paralyzed' as well. On the other hand, without a period of pre-amputation paralysis, the emulator is never confronted with anything to contradict its prior input-output mapping (to see this point, the difference between (i) a lack of information and (ii) information to the effect that nothing happened must be kept in mind). On the assumption that phantom limbs are the result of the operation of an emulator, we have a possible explanation for this phenomenon.

Regardless of whether a motor plan is conceived as a dynamic plan or a kinematic plan, it should be clear from the above example that a plan is one thing, and the sequence of proprioceptive sensations produced when it is executed is another. The simulation theorist would have to maintain that those who have an inability to produce motor imagery of a certain sort (because their phantom is paralyzed) also have an inability to produce motor plans of a certain sort. But subjects with paralyzed phantoms can obviously make motor plans – they know all too well what it is that they cannot do, they cannot move their limb from their side to straight in front, for example. The plan is fine. What is wrong is that when the plan is executed on the phantom, nothing happens: the proprioceptive signal remains stubbornly static. A motor plan is one thing, a sequence of proprioception is another. The simulation theorist conflates them. In any case, this issue will be revisited in the case of visual imagery, where the difference between a motor plan and the resulting sensations is more clear.

Beside the simulation and emulation theories, there are two other possible accounts of motor imagery. First, it might be thought that motor imagery is just the memory of the feelings of previous overt actions. While it is of course true that one can remember such past experiences, the hypothesis that this is what motor imagery consists of has nothing in its favor. Brain areas known to subserve memory have not been found to be notably active during motor imagery. Also, there is no reason to think that the phantom limb patients with paralyzed phantoms have an inability to remember moving their actual limb.

The other potential account is that just as visual imagery is (perhaps) the result of 'top-down' activation of cortical areas involved in overt visual perception, perhaps motor imagery is the product of the 'top-down' activation of areas normally involved in proprioception. This view is more or less the polar opposite of the simulation theory. The simulation theory takes motor imagery to be the product of the covert operation of the *efferent* areas, where as this

view takes it to be the covert top-down activation of the *afferent* areas. But note that this view is *not* incompatible with the emulation theory. In fact, the emulation theory is a specific theory about just how these afferent areas might be driven in a top-down manner – a view of how they can be driven by efferent areas, and what this means. Recall that according to the KF-control scheme, the sensory input is *processed* by the emulator, in that the raw sensory input is combined with the emulator's expectation in order to produce the final percept, and this final percept is encoded by updating the emulator's state estimate. It is the emulator's output that is used as the optimal perceptual signal. That is, during perception the emulator *is* an afferent area, and the KF-control scheme is a specific view as to what it means for an afferent area of this sort to be driven by top-down processes. On the emulation theory of motor imagery, motor imagery involves the operation of both efferent and afferent areas: the efferent areas (controller) are what is driving the afferent areas (emulator). This will become more clear in the case of the next topic, visual imagery.

3. Visual Imagery

So far the only applications of the strategy of emulation have been to motor control, proprioception, and motor imagery. But the same basic scheme is obviously applicable to other modalities, provided that they have some relevantly exploitable structure. Section 3.1 introduces two models by Bartlett Mel (Mel, 1986; Mel, 1988), robots that can use imagery to solve certain problems. Though Mel does not describe in this way at all, they generate imagery by operating emulators of the motor-visual loop. The details of how these systems generate imagery is no doubt distinct from how nervous systems do it, but at a gross level there is evidence that the same basic architecture is in play. So after getting the basic idea across with Mel's models, I will turn in sections 3.2 and 3.3 to providing evidence that visual imagery is generated via the operation of a motor-visual emulator in a way at least roughly analogous to that suggested by the models. In section 3.4 I introduce a distinction between modal and amodal imagery. Modal imagery, as exemplified in Mel's models, is imagery based on the operation of an emulator of the sensory system itself, whereas amodal imagery is based on the operation of an emulator of the organism and its environment: something like solid objects in egocentric space. I show how the two forms of emulation can work in tandem.

3.1 Murphy

Murphy (Mel, 1988) is a robot whose job is to move its arm while avoiding contact with obstacles so that its hand can grasp objects. The arm has three joints – a shoulder, elbow and wrist – all of whose motion is confined to a single plane. There is a video camera trained on the arm and workspace which drives a 64 x 64 grid of units, each effectively a 'pixel' of an image of the workspace. Murphy controls the limb on the basis of the image projected on the grid, where the arm, target, and obstacles are all clearly visible. Murphy operates in two stages. In the first stage, Murphy simply moves its arm around the workspace until it manages to find a path of movement that gets its hand to the target without impacting any obstacles. Because the arm has redundant degrees of freedom, it is not a trivial problem to find a path to the target. Often what looks initially like a promising route ends up being impossible to manage, and Murphy must backtrack, attempting to work its limb around obstacles in some other way.

The twist on this is that each unit in the visual grid is actually a connectionist unit that receives an input not only from video camera, as described, but also receives a copy of Murphy's motor command (e.g. increase elbow angle, decrease sholder angle), as well as inputs from neighboring units. During the first phase just described, while Murphy is overtly moving its limb around, the units on the grid are learning the *forward mapping* of the motor visual loop. That is, the grid learns that if the visual input at t_1 is x_1, and motor command m_1 is issued, the next visual input, at t_2, will be x_2. Qualitatively, Murphy's overt motor-visual system is a plant, implementing a forward mapping from current states and motor commands to future states. The visual grid units monitor the inputs to this system (the motor commands) and see what outputs the system produces on their basis (in this case, system's outputs are patterns of activations on the visual grid).

After a certain amount of experience solving movement problems overtly by trial and error, Murphy gains a new ability. When the visual grid has learned the forward mapping, Murphy is able to solve the problems off-line using visual imagery. It gets an initial visual input of the workspace, including the configuration of the arm and location of target and obstacles. It then takes the real arms and camera off-line, and manipulates the visual grid with efferent copies. It thereby moves the image of its arm around, seeing what sequences of movement impact objects, sometimes backing up to try another potential solution, until it

finds a path that works. At that point, it puts the real arm and camera back on line an implements, in one go, the solution.

Mel nowhere puts any of this in terms of control theory or forward mappings, etc. Rather, he describes it simply as a connectionist network that learns to solve problems through imagery. Nevertheless, during the imagery phase it is a clear that the connectionist network is implementing a pseudo-closed-loop control scheme. The grid itself actually serves double duty as both the medium of visual input, and the emulator of the motor visual loop. When operating on-line, the grid is driven by the video camera. When off-line, it is driven by the activity of its own units and the motor inputs. Because the grid is used for both, the system at least has a capacity that it never in fact uses. Specifically, it never operates in anything like a Kalman filter mode. This would involve having the imagery capacity engaged during overt operation. On such a mode of operation, the grid would always form an expectation of what the next visual input would be on the basis of the current visual representation and the current motor command. This expectation would then take the form of some degree of activity in some of the units anticipating activation from the camera. This would be helpful in cases where the video input was degraded, and forming an anticipation might be crucial in interpreting the input.

The next model, also by Mel, is similar to Murphy.[2] It consists of a robot with two video cameras, to provide ocular disparity, each of which drives an input grid similar to Murphy's. This robot has no limbs, but rather moves itself around wire frame objects. For example, it might move towards or away from a wire frame cube, or circle around it. And just as with Murphy, there are two modes of operation. There is an initial overt mode, during which the robot moves around various wire frame objects. All the time, the units of the visual grid are getting activation not only from the video cameras, but from efferent copies and also from connections to other units in both grids. Again, the grids learn the forward mapping of the motor-visual loop. Once this is complete, the robot is able to engage in visual imagery in which it can mentally rotate, zoom, and pan images, including images of novel shapes. Upon receiving an initial image on both visual grids from some object, the system takes its motor system and video cameras off-line, and drives the visual grid with efferent copies. It can mentally rotate the image by issuing a command that would normally circle the object. It can zoom into or out of the image by issuing a command the would (overtly) move towards or away from the object.

---

[2] One difference is that Murphy was an actual robot, whereas this model discussed here is completely virtual.

Again, Mel does not couch any of this in terms of control loops, emulators, etc. And again, the potential for exploiting the grids as part of a Kalman-filer-like mechanisms for processing perception is not explored.

3.2 Visual imagery and visual areas of the brain.

On the KF-control scheme the emulator is a system that processes sensory information, specifically, it produces an expectation, and combines it with the sensory residual in order to yield a best estimate of the state of the observed process. For now the details of the perceptual situation is not the focus. Rather the point is merely that the emulator is involved both in imagery and in perceptual processing. Mel's models are concrete examples of systems in which the emulator does double-duty, even though in Mel's models the emulators never do both simultaneously.

The hypothesis that this scheme is used by real nervous systems makes a number of predictions, the first of which is that visual 'perceptual' areas will be active during visual imagery. And indeed there is much evidence not only that such areas area active, but that their activity is selectively similar to the activity of such areas during the analogous overt perceptual situations. Since the focus is currently on imagery that is modality specific (see section 3.4), the relevant visual areas will include early visual areas.

A number of researchers have reported finding activity in primary visual areas during visual imagery (Chen et al., 1998; for a recent review, see Behrmann, 2000). Kosslyn et al. (1993, 1995) found that visual imagery not only activates primary visual cortex, but that imagining large objects activates more of this area than does imagining smaller objects, indicating that it is not only active as during imagery, but that details about the kind of activity is presents is also parallel.

In an extremely suggestive study, Martha Farah (Farah et al. 1992) reports on a subject who was to have a unilateral occipital lobectomy. The subject was given a number of imagery tasks before the operation, including tasks in which she was asked to imagine moving towards objects (such as the side of a bus) until she was so close that the ends of those objects were at the edge of her imagined 'visual field'. After the removal of one of her

occipital lobes, the subject was re-tested, and it was found that the distance increased. This suggests that, much like in the case of Mel's models, the image is actually produced on a topographically organized medium, and manipulated via efferent copies. With a smaller screen, 'walking towards' an imagined object reaches the point where the edges of the object are at the edges of the topographic medium at a greater distance than with a larger screen.

3.3 Imagery and motor control

As Mel's models suggest, some kinds of visual imagery might, surprisingly, require the covert operation of motor areas. In this section I will point out some evidence indicating that in fact motor areas are not only active during, but crucial to, certain sorts of visual imagery. Activity in premotor areas has been widely shown to occur during imagery tasks requiring image rotation. Richter et al. (2000) demonstrated this with time-resolved fMRI, a result confirmed by Lamm et al (2001).

While such studies are interesting, the theory I am articulating here makes predictions more detailed than the simple prediction that motor areas are active during visual imagery. It makes the specific prediction that they are active in producing motor commands of the sort that would lead to the overt counterpart of the imagined event. Enter a set of experiments done by Mark Wexler (Wexler et al. 1998) in which subjects were engaged in imagery tasks while simultaneously producing an overt motor command. The experiment was suggested by the observation that during some imagery tasks, subjects with specific cortical lesions would try to reach out with their hands to rotate the image on the computer display. This suggested that perhaps visual imagery in normal subjects involved the covert execution of an overt movement – in this case reaching out to turn some object.

To test this, Wexler et al. designed experiments in which subjects had to solve problems already known to involve certain kinds of visual imagery, specifically the rotation of visually presented shapes. At the same time, subjects were to hold and apply a torque (twisting force) to a handle. Results showed that when the required direction of image rotation and the actual applied torque were in the same direction, performance was much better than trials in which the direction was different. The natural interpretation of this data is that is shows not only that motor areas are involved in visual imagery, but that their involvement

takes the form of producing those specific motor commands that would be required to produce the overt movement corresponding to the relevant image transformation.

As an aside, it is obvious that a 'simulation' theory of visual imagery parallel to the simulation theory of motor imagery discussed in Section 2, according to which visual imagery is the product merely of a covert motor plan, would be a non-starter. While motor areas are involved, clearly the motor plan by itself underdetermines the nature of the imagery. Presumably imagining twisting a 'd' and a 'b' involve identical motor plans – twisting the grasping hand left or right. But the nature of the image produced is quite different, as it would have to be to solve the problem. The difference in this case is that *the states of the emulators in the two cases are different*, and so driving them with the same motor command does not yield the same result. One yields a rotated 'd', the other a rotated 'b'. As I tried to show in section 2, the same under-determination is present in the case of motor imagery, even though a number of factors can obscure this fact.

3.4 Modal imagery vs. amodal imagery

An emulator is an entity that mimics the input-output function of some other system. But even when the same control loop is involved, different systems might be being emulated. In Mel's model, for example, the elements emulated are the pixels on the visual input grid, and the relevant dynamics concerns the way in which one pattern of active pixels plus a motor command leads to a new pattern of active pixels. Nowhere does the emulator have a component that corresponds to the arm, or hand, or elbow angle. A given active pixel might correspond to part of the hand at one time, and an obstacle at another time. This is of course an entirely legitimate use of emulation to produce imagery, in this case specifically visual imagery. The visual input grid is a system of states, and there are rules governing the transitions from one state to the next.

But the emulation might also take a different form. In this example, it might take the form of an emulator with components corresponding to parameters of the arm itself. This system would learn how these parameters change as a function of the motor command sent to them – hence the forward mapping learned would not be (current pixel pattern + current motor command) -> (next pixel pattern), but rather (current arm parameters + current

motor command) -> (next arm parameters). The system would then subject the arm emulator's state to a 'measurement' that would capture the way in which the real video camera maps arm states to grid images.

How can two different systems be legitimate objects of emulation like this? As I mentioned, the visual grid is a set of elements whose states change over time in predictable ways. Given visual grid pattern p_1, and motor command m_1, the next visual input pattern will be p_2. This is a forward mapping that can be emulated. But behind the scenes of this visual grid system is another system, the system consisting of the arm, workspace, video camera, and so forth. This system also consists of entities whose states interact and change over time according to laws – in this case the laws of mechanics and dynamics. And as such, it too implements a forward mapping that can be emulated. And it is obviously true that the visual grid system has the rules of evolution that it has because of the nature of the arm/workspace system and its laws of evolution. If the arm were heavier, or made of rubber rather than steel, then there would be a different mapping from visual input grid patterns plus motor commands to future visual input patterns. Which system is being emulated is determined by the number and nature of the state variables making up the emulator, and the laws governing their evolution. Mel's Murphy for example uses an emulator whose states (levels of activation of pixel units) obviously correspond to the elements of the visual input grid.

Either way, the end result is a system that can produce visual imagery. But they do it in rather different ways. The Mel type systems produce it by constructing an emulator of the sensory input grid itself. In this case, the emulator's state just is the visual image, there is no measurement, or if you like, the measurement consists of the identity matrix. The other system I described produces visual imagery by constructing and maintaining an emulator of the *arm's* state, and then subjecting this to a 'visual measurement', similar to the measurement that the video camera subjects the real robotic arm to, in order to produce a mock visual image.

Both ways are shown in Figure 7, in which three control loops are represented. The top, boxed in red, is just the actual system. The process is the organism and its environment, the process changes state as a function both of its own dynamic as well as motor commands issued by the organism. The organism's sense organs produce a measurement of

the state of this, resulting, in the visual case, in a topographic image on the retina or primary visual cortex. Nothing new here.

The second, boxed in green, corresponds to a modality-specific emulator, as exemplified in Mel's models. This emulator's states just are states corresponding to elements in the topographic image. So long as the elements in the image itself compose a system whose states are to some degree predictable on the basis of previous states and motor commands, it implements a forward mapping that can be emulated. Since the emulated system just is the visual input medium, no measurement is needed in order to arrive at a visual image.

The third, boxed in blue at the bottom, represents an emulator of the organism and its environment, and for simplicity we can assume that this consists of a sort of model of solid objects of various sizes and shapes in the organism's egocentric space (this will be discussed more in section 4).
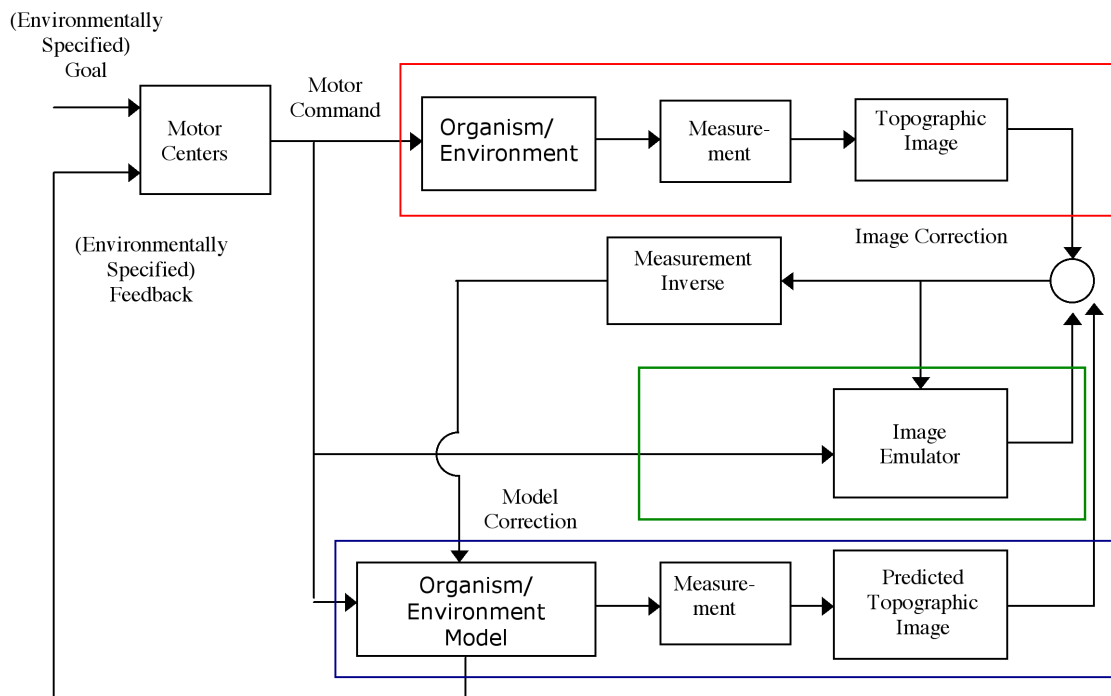

Figure 7. A KF-Control scheme using two emulators: one a modality-specific 'image' emulator; the other an amodal 'spatial' emulator.

To run through an example, suppose that we have an organism with a visual sensory modality standing directly in front of a white cube, and it moves to the right. We can describe what happens in the red box by pointing out that a white cube is 2 meters in front of the organism, which is causing a square retinal pattern of stimulation. The creature moves to the right, and so the cube is now slightly to the creature's left, and the new retinal image (produced by the 'measurement') is a square pattern of stimulation on the retina, but slightly shifted to the right from where it was before the movement.

The green-boxed material represents a Mel-type sensory emulator. This system consists of a grid of pixels corresponding to the visual input. Initially it has a square pattern of activation. It processes a 'move right' efferent copy by altering the pattern of active pixels according to something like the following rules: A pixel p is active at t_2 if and only if the pixel to its immediate left was active at t_1. This will have the effect of sliding the image one pixel to the right. Since the emulator in this case just consists of nothing but the topographic image, no measurement is needed (or if you like, the measurement consists of the identity matrix). If this is being operated off-line, the resulting image is just visual imagery. If it is on-line, the resulting image is an *a priori* estimate of the 'observed' visual input, which will be combined with the sensory residual to yield the final state estimate.

What about the blue box? Here there is an inner model that represents the organism's egocentric environment: in this case a cube directly in front of the organism. Subjecting this to a visual 'measurement' yields square topographic image. The 'move right' efferent copy alters the state of the model, so that the object is now represented as being in front and slightly to the left of the organism. Subjecting this to a 'visual measurement' yields a new topographic input image, similar to the previous one, only with the patterns altered slightly. If operated off-line, the result would be visual imagery, if measured, or amodal imagery, if not subjected to a measurement. If operated on-line, the state of this model would constitute the *a priori* estimate of the layout of the organism's environment, to be modified to some extent by the sensory residual, if any, to yield the final state estimate of the organism's environment.

The two methods are not incompatible, and we could easily imagine a system that uses both, as in Figure 7. This system would run two emulators, one of the sensory periphery as in Mel's models, and also an emulator of the organism/environment, as described above.

This system would have not one but two *a priori* estimates, which could be combined with each other and the sensory residual in order to update both emulators.

An amodal emulator (in this example, the organism/environment model) supplies a number of advantages, stemming from the fact that its operation is not tied to the sensory modality. First, the organism/environment model can easily represent states that are not manifest to the sensory periphery. For example, it can easily represent an object as moving from the front, around to the left, behind, and then up from the right hand side as the organism turns around, or objects behind opaque surfaces. This is not something that could easily be done with a Mel-type system. In a system that includes both a modal and amodal emulator, the amodal emulator could provide an anticipation in such cases, such as that the leading edge of a square will appear in the visual periphery as the spinning continues.

Second, the same amodal emulator model might be used with more than one modality-specific emulator. I won't bother with a diagram, which would be rather messy, but the idea is that an organism that has, say, visual and auditory modality specific emulators might be able to run both in tandem with an amodal emulator. In such a case the amodal emulator would be subject to two different 'measurements', a visual measurement, yielding an expectation of what should be *seen* given the current estimate of the state of the environment, and an auditory measurement yielding an expectation of what should be *heard* given the current estimate of the state of the environment. And the amodal emulator would be updated by both sensory residuals, resulting in a state estimate that effectively integrates information from all modalities as well as a priori estimates of the state of the environment (van Pabst and Krekel, 1993; Alain et al., 2001).

There are additional possibilities that could be explored here. For now I just want to point out that the scheme I am articulating here allows for (at least) two very different kinds of imagery: modality specific imagery, which is the result of running an emulator of the sensory modality off-line (as in Mel's models); and amodal imagery, which results in the off-line operation of an emulator of the plant (process) without a corresponding sensory-specific measurement. Such amodal imagery might be accompanied with modality-specific imagery, but it might not. More will be said about such cases in Sections 4 and 5.

3.5 Discussion

There are a number of aspects of visual imagery that have not been covered in the discussion of this section. For example, I have said nothing about how a system decides when using imagery is appropriate. I have not mentioned anything about how the imagination process gets started – in Mel's models they begin with an initial sensory input that is subsequently manipulated via imagery, but clearly we can engage in imagery without needing it to be seeded via overt sensation each time. Furthermore, many sorts of visual imagery don't obviously involve any sort of motor component, as when one simply imagines a vase of flowers.

As far as the first of these points go, they are correct. An emulator by itself does not decide when it gets operated on-line vs. off-line. Presumably there is some executive process that makes use of emulators, sometimes for imagery, sometimes for perceptual purposes. I am not making the outrageous claim that the brain is *nothing but* a big KF-control system. Of course other processes are required. But they would be required on any other account of imagery and perception as well.

Now once it has been decided that the emulator should be run off-line, it is presumably up to some other system to seed the emulator appropriately. Again, this is a process outside the scope of the present focus. The initial state of the emulator gets set somehow, perhaps a memory of a state it was once in. Again, a KF-control architecture is necessarily part of a larger cognitive system that includes memory and executive processes. My theory is that when this cognitive system is engaged in imagery, it is exploiting an emulator, and when perceiving (next section) it is using emulators as part of a KF-control scheme. The fact that there connections to broader cognitive components not a weakness of my account, but rather a necessary feature of any account of imagery and perception. Detailing such connections would be one of many tasks required in order to completely fill out the account I am outlining in this article.

On the theory I am pushing, visual imagery is 'mock' input generated from the operation of an internal emulator. The imagery thus produced depends on what sequence of states the emulator goes through. And this depends on at least three factors. The first, which I will mention and drop, is whatever executive process there is that can initially configure the emulator, as mentioned in the previous paragraph. The second is the emulator's own internal dynamic; depending on what is being emulated, the state might or might not evolve over time on its own in some specific way. The third factor is efferent copies of motor

commands. In this section I have focused on imagery produced through the emulation of processes that have no or minimal dynamic of their own, but depend for their evolution on being driven by efferent copies. Mel's models highlight this, as does the imagery involved in Wexler's studies. But this has been no more than a focus of my discussion, not a necessary feature of the model. Some bouts of imagery might involve configuring the emulator to emulate a static process, such as looking at a vase of flowers, where nothing changes over time. In this case, there would be neither any emulator-internal, nor efferent copy-driven dynamic to the emulator's state. It would be constant, and yield the more-or-less constant mock sensory input of a vase of flowers. In other cases, there might be a dynamic driven by the emulator's state, as when I imagine pool balls hitting each other. In this case, the imagined scene is dynamic, but the dynamic is not driven by any efferent copies, but by the modeling of processes that evolve over time on their own. The model thus includes these other sorts of imagery as special cases. I focus on the case involving efferent copies to bring out the nature of the fullest form of the model.

Furthermore, the ability of the scheme to handle both modal and amodal imagery surely allows for explanations of various imagery phenomena. Some sorts of imagery are more purely visual than spatial, as when you simply imagine the colors and shapes of a vase of flowers. Such imagery need not involve imagining the vase of flowers as being anywhere in particular, and might be something like the operation of a purely modal, visual, emulator. There is a difference between this sort of case, and a case where you imagine a vase of flowers sitting on the desk in front of you. In this case, the imagined vase not only has its own colors and textures, but it is located in egocentric space – you might decide where a vase should be placed so as to obscure a picture on the desk on the basis of such imagery, for example. This might involve both the modal and amodal emulators. And some tasks requiring spatial imagery might not involve any notably visual component at all.[3] Differing intuitions about whether or not imagery is involved in this or that case might be the result of thinking of different kinds of imagery, kinds that can all be described and explained in the current framework.

A final comment before moving on. The present theory offers a single framework within which both motor and visual imagery can be understood. This is remarkable in itself, since

---

[3] For example: If x is between a and b, and y is between x and z, is x necessarily between a and b? There is reason to think that such questions are answered by engaging in spatial imagery, but little reason to think that much in the way of specifically visual mock experience is involved, though of course it might be.

surprisingly the dominant accounts of motor and visual imagery in the literature are not at all parallel. Dominant ideas concerning motor imagery, as we have seen, equate it with the covert operation of *efferent* processes – either a mere efferent copy or a completed motor plan. Dominant ideas about visual imagery treat is as the covert 'top-down' stimulation of *afferent* areas. Given that motor and visual imagery are both imagery, the fact that these two dominant explanations seem so *prima facie* different is at least surprising, at most embarrassing.

The theory defended here unifies both accounts seamlessly. Imagery is the result of the off-line operation of emulators. In the motor case, such emulators are predominantly driven by efferent copies, especially all cases of motor imagery studied where subjects are invariably asked to imagine engaging in some motor activity. Hence the ubiquitous involvement of motor areas in motor imagery. Visual imagery also involves the off-line operation of an emulator (in this case, a motor-visual emulator). But in many cases the motor aspect is minimal or absent, since the emulation required to support the imagery does not require efferent copies, though of course in some cases it does (as in Wexler's studies).

4. Perception

4.1. Sensation vs. perception

Psychologists and philosophers have often distinguished between sensation and perception. The distinction is not easy to define rigorously, but the general idea is clear enough. Sensation is raw sensory input, while perception is a representation of how things are in the environment based on, or suggested by, this input. So for example when looking at a wire-frame cube, the *sensory* input is twelve line segments; four horizontal, four vertical, and four diagonal, arranged in the familiar way. What one *perceives* is a cube, a three-dimensional object in space. That the perception is an interpretation of the sensory input is highlighted by the fact that one can, at least in some cases, switch which face of the cube is in front, as with the Necker cube. Here there are two different interpretations that can be placed on the same sensory input; two different perceptual states based on the same sensory state.

The sorts of representational states that result from perception are extremely complex, but for purposes of the present discussion I will focus on what I take to be the core aspects. Through perception we become aware of objects in our surroundings. A bit more specifically, we become aware of some number of objects, their rough sizes and shapes, their dynamical properties (especially movements), and their egocentric locations. To have some handy terminology, I will refer to this as the *egocentric space/object environment* or ESOE. Clearly, one of the primary functions of perception is the formation an accurate representation of the ESOE.

Look again at figure 7. In Section 3 I highlighted one aspect of this diagram – its combination of modal and amodal emulators. But now I want to draw attention to another aspect, which is that the feedback from the emulator to the controller does not go through the measurement process. In Figure 2, the control context within which we started involved a controller that was given a goal state, and got feedback that was used to assess the success of the motor program in achieving that goal state. On the feedback control scheme, the feedback is necessarily whatever signal is produced by the plant's sensors, and this imposed as a requirement that the goal specification given to the controller be in the same format as the feedback, for only if this is the case can an assessment between the desired and actual state of the plant be made. That is, the goal state specification had to be in sensory terms.

On the pseudo-closed-loop scheme of Figure 4, and the KF-control scheme of Figure 6, the idea that the feedback sent from the emulator to the controller was also in this 'sensory' format was retained. In the latter case this was made explicit by including a 'measurement' of the emulator's state parallel to the measurement of the real process in order to produce a signal in the same format as the real signal from the plant.

But retaining this 'measurement' is not desirable in many cases. The real process/plant has many state variables, only a sampling of which are actually measured. In the biological case, access to the body's and environment's states through sensation is limited by the contingencies of the physiology of the sensors. A system with an emulator that is maintaining an optimal estimate of all the body's or environment's relevant states is needlessly throwing away a great deal of information by using only the mock 'sensory' signal that can be had by subjecting this emulator to the same measurement process. And there is no need to do this, either. The emulator is a neural system, any and all of its states

can be directly measured. This is the meaning of the fact that in Figure 7 the feedback to the controller comes directly from the emulator, without the modality-specific 'measurement' being made.

The practical difference between the two cases is not insignificant, since, as already mentioned, the measurement process might very well throw out a great deal of useful information. But the conceptual difference is more important for present purposes. It is not inaccurate to describe the 'measured' or 'modal' control schemes, including the KF-control scheme of Figure 6, as systems that control *sensation*. Their goal is a sensory goal, they want their sensory input to be like thus-and-so, and they send control signals out that manage to alter their sensory input until it is like thus-and-so. The information they are getting is exclusively information about the state of the sensors. But on the unmeasured amodal variant, the controller has its goal specified in terms of objects and states in the environment, and the feedback it gets is information about the objects in its egocentric environment.

The less sophisticated systems are engaged with their sensors. And this is true both on the efferent and afferent ends. The more sophisticated systems have their goals set in terms of objects and locations in the environment, and get information in terms of objects and locations in their environment.

4.2. The egocentric space/object emulator

If the relevant emulator for perception were an emulator of the sensory surface, as in Mel's models, then there would be little question concerning their states – they are just the states of the components of the sensory organs just as the units in Mel's simulations are pixels of a visual image. But I have claimed that perception involves the maintenance of an emulator whose states correspond to states of the emulated system. This can be made sense of readily in the case of proprioception and the MSS, as done above. The relevant states are the dynamic variables of the MSS. But what about other sorts of perception, such as visual perception? What is the emulated system, and what are its states, if not the sensor sheets? To a plausible first approximation the emulated system is the organism and its immediate environment; specifically, objects of various sizes, shapes, and egocentrically specified locations, entering into force-dynamic interactions with each other and the organism.

This involves a combination of the 'where', 'what' and 'which' systems. The 'what' and 'where' systems are posited to be located in the ventral and dorsal processing streams respectively (Ungerleider and Haxby, 1994). The ventral 'what' stream proceeds from early visual areas in the direction of the temporal lobes, and appears to be concerned with identifying the type of object(s) in the visual field and their properties. The dorsal 'where' stream proceeds from early visual areas to the parietal areas, and is primarily concerned with the location of objects. In addition to these, a 'which' system, whose task is indexing and tracking object identity, also appears to be in play (Pylyshyn, 2001; Yantis, 1992).

These systems comprise the core of the ESOE emulator on the present account. During perception they jointly maintain an estimate of the relevant layout of the environment, especially the number, kind, and egocentric locations of objects. Anticipations about how this layout will change are continually produced, both on the basis of the organism's own movements that result in changes in the egocentric location of objects, as well as anticipated changes brought about by the dynamics of the objects themselves (hence identifying the kind of object in question is crucial). This estimate provides a framework for interpreting sensory input, and is subject to modification on the basis of sensory information.

4.3. Kosslyn on perception and imagery

Stephen Kosslyn has developed one of the most influential accounts of the nature of visual perception and its relation to visual imagery. In this section I will merely outline the relations between his account and mine.

On both accounts, perception and imagery are closely related. On Kosslyn's view, imagery is produced when areas involved in perceptual processing are activated by 'top-down' influences. In some cases of what Kosslyn calls 'motion added' imagery, he maintains that this top-down influence takes the form of the influence of covert motor processes. By the same token, Kosslyn maintains that imagery processes are used to aid perceptual processing, by filling in missing information on the basis of expectations, for instance. As Kosslyn and Sussman (1995) put it, the view is "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."

Clearly all aspects of this are not only embraced on the account I am here articulating, but they are given an explanation. The differences are only slight differences of emphasis. On my account, imagery and perception are not two separate processes, such that one can aid the other. Rather, they are two modes of operation of the same process -- the continual updating of an estimate of the state of the environment and primary perceptual areas. During perception, this process is influenced (though not determined) by sensory information, during imagery it is not. The account here also reveals more conspicuously the relation of imagery and perception to motor processes. It also provides a framework within which imagery and perception in other modalities, including motor imagery, can be naturally fit.

4.4. Discussion

In a sense, what I have said about perception glosses almost entirely over what most researchers take to be most important. A standard and unobjectionable view of what perception involves is that it is the creation of a representation of the organism's environment's layout from bare sensation. On my account, a good deal of this is embedded in the part of the system that goes from a sensory residual to an *a posteriori* correction. The key here is the 'measurement inverse', which is just a process that takes as input sensory information, and provides as output information in terms of the states of the emulator. In the case of the amodal ESOE emulator, this process goes from sensory signals to information about the layout of the environment. This is the process that is the paradigmatic perceptual process, and I say next to nothing about it, except to locate it in the broader framework.

But to fixate on this is to miss the import of the scheme I am articulating. The point is to show how *this* process is part of a larger process, and does so in such a way as to hopefully highlight two related points. The first is the large-scale nature of perception, the second is the fact that perception in one aspect of a complicated process that intimately involves motor control and imagery. I will address these in reverse order.

The standard view of perception says nothing at all about how or even if perception has any connection at all to systems involved in motor control, imagery, or cognition, and in fact few of the proposals one finds concerning the mechanisms of perception draw any such connections. The present account, by contrast, argues that the brain engages in a certain very flexible and powerful sort of information processing strategy, one that simultaneously addresses all of these (and perhaps makes others possible as well – see Section 5). This seems plausible, for surely to treat perception, imagery and motor control as functionally distinct modules is to significantly distort the phenomena.

This leads to the second point, which is that the current scheme, exactly because it treats perception as one aspect of an integrated information processing strategy, sheds light on the nature of perception itself. In the first place, the scheme highlights the extent to which the outcome of the perceptual process, the state estimate embodied in the emulator, is tuned to sensorimotor requirements. The emulator represents objects and the environment *as things engaged with* in certain ways as opposed to how they are considered apart from their role in the organism's environmental engagements. The perceived environment is the environment as made manifest through the organism's engagements, because the emulator that supplies the perceptual interpretation is an emulator of the agent/environment interactions.

Another shift in emphasis suggested by this account is that perception is shown to be not a matter of starting with materials provided in sensation and filling in blanks until a complete percept is available. Rather, complete percepts of the environment are the starting point, in that the emulator always has a complete and potentially self-contained ESOE estimate up and running. This self-contained estimate is operational not only during imagery, but presumably also during dreaming (see Llinas and Pare, 1991). The role played by sensation is to constrain the configuration and evolution of this representation. In motto form, *perception is a controlled hallucination process.*[4]

---

[4] I owe this phrase to Ramesh Jain, who produced it during a talk at UCSD.

5. General discussion and conclusion

5.1 Perception and imagery

The imagery debate, well known to cognitive neuroscientists, is a debate concerning the sort of representations used to solve certain kinds of tasks. The two formats under consideration are *propositions* and *images*. As is often the case, definitions are difficult, but the rough idea is easy enough. Propositions are conceived primarily on analogy with sentences, and images on analogy with pictures. In its clearest form, a proposition is a structured representation, with structural elements corresponding to singular terms (the content of which prototypically concerns objects) and predicates (the content of which prototypically concerns properties and relations), as well as others. This structure permits logical relations such as entailment to obtain between representations. On a caricature of the pro-proposition view, perception is a matter of turning input at the sensory transducers into structured language-like representations; cognition is a matter of manipulating such structured representations in order to draw conclusions in accord with laws of inference and probability.

By contrast, images are understood as something like a picture: a pseudo-sensory presentation similar to what one would enjoy while perceiving the depicted event or process. Perception is a matter of the production of such images. Cognition is a matter of manipulating them.

According to the present theory, one of the central forms of imagery is amodal spatial imagery. It will often be the case that this imagery is accompanied by modality specific imagery, for the same efferent copies will drive both the modality specific emulators as well as the amodal spatial emulator. Indeed, the fact that there are in-principle isolatable (see Farah et al., 1988) aspects to this imagery may not be introspectively apparent, thus yielding the potentially false intuition that 'imagery' is univocal.

Amodal spatial imagery is not a clear case of 'imagery' as understood by either the pro-proposition or pro-imagery camps; nor is it clear that such representations are best conceived as propositions. Like propositions, this imagery is structured, consisting at least of *objects* with *properties*, standing in spatial and dynamical relations to each other (Schwartz, 1999). They are constructs compositionally derived from components that can

be combined and recombined in systematic ways. An element in the model is an object with certain properties, such as location and motion, and this is analogous in some respects to a proposition typically thought of as the predication of a property to some object.

Amodal imagery is emphatically *unlike* a picture. Rather, modal imagery is the sort of imagery best described as picture-like. The distinction is difficult in part because we typically, automatically and unconsciously, interpret pictures as having spatial/object import. But strictly speaking this import is not part of the picture. Similarly, bare modal imagery is unstructured, lacks any object/spatial import. But because of the potentially close ties between modal and amodal imagery, modal imagery is typically, automatically and unconsciously, given an interpretation in terms provided by amodal imagery. The point is that amodal imagery is not picture-like.

On the other hand, amodal spatial imagery *is* a representation of the same format as that whose formation constitutes perception, for the simple reason that perception just is, on my account, sensation given an interpretation in terms of the amodal ESOE emulator. Thus although amodal imagery is not picture-like, it is also not obviously sentential either. These amodal environment emulators are closely tied to the organism's sensorimotor engagement with its environment. The model is driven by efferent copies, and transformations from one representational state to another follows the laws of the dynamics of movement and engagement, not of logic and entailment (as typically understood), or at least not *only* according to logic and entailment. Unlike a set of sentences or propositions, the amodal environment emulator is spatially (and temporally) organized.

I don't have any answers here. I mean merely to point out that if in fact amodal object/space imagery is a core form of neurocognitive representation, then this might go some way to explaining why two camps, one insisting one understanding representation in terms of logically structured propositions, and the other in terms of picture-like images, could find themselves in such a pickle. The camps would be trapped by the two dominant metaphors for representations we have: pictures and sentences. I am suggesting that neither of these metaphors does a very good job of capturing the distinctive character of amodal imagery, and that if progress is to be made, we might need to abandon these two relic metaphors, and explore some new options, one of which I am providing.

5.2 Cognition

Kenneth Craik (1943) argued that cognition was a matter of the operation of small scale models of reality represented neurally in order to anticipate the consequences of actions, and more generally to evaluate counterfactuals. Phillip Johnson-Laird has refined and developed this approach under the title of 'Mental Models', and it is currently a dominant theory of cognition in cognitive science. Johnson-Laird describes mental models as representations of "spatial relations, events and processes, and the operations of complex systems", and hypothesizes that they "might originally have evolved as the ultimate output of perceptual processes" (Johnson-Laird, 2001). The representations embodied in the amodal ESOE emulators are of exactly this sort.

Johnson-Laird's mental models, while arguably based on something like the representations made available through such emulators, involve more than I have so far introduced. Specifically, on Johnson-Laird's account they are manipulated by a system capable of drawing deductive and inductive inferences from them. The difference between a mental models account and an account that takes reasoning to be a matter of the manipulation of sentential representations according to rules of deduction and probability is thus not that logical relations are not involved, but rather that the sort of representation over which they operate is not sentential, but spatial/object model. Exactly what is involved in a system capable of manipulating models of this sort such as to yield inferences is not anything that I care to speculate on now. I merely want to point out that the individual mental models themselves, as Johnson-Laird understands them, appear to be space/object emulators, as understood in the current framework.

In a similar vein, Lawrence Barsalou (1999; Barsalou et al, 1999) has tried to show that what he calls 'simulators' are capable of supporting the sort of conceptual capacities taken to be the hallmark of cognition. Barsalou's simulators are capacities for imagistic simulation derived from perceptual experience. He argues that once learned, these simulators can be recombined to produce 'simulations' of various scenarios, and that such simulations subserve not only cognition, but serve as the semantic import of linguistic expressions.

I am not here specifically endorsing either Johnson-Laird's or Barsalou's accounts, though I do think that they are largely compatible, and each has a lot going for it. My point is merely to gesture in the direction in which the basic sort of representational capacities I have argued for in this article can be extended to account for core cognitive abilities.

5.3 Other applications

There are a great number of other potential applications of the KF-control framework in the cognitive and behavioral sciences. I will here touch on just a few.

Antonio Damasio (1994) has argued that skill in practical decision making depends on emotional and ultimately visceral feedback concerning the consequences of possible actions. The idea is that through experience with various actions and the emotionally charged consequences that actually follow upon them, an association is learned such that we tend to avoid actions that are associated with negative emotions or visceral reactions. The relevant part of his theory is that it posits an 'as-if loop', based in the amygdala and hypothalamus, that learns to mimic the responses of the actual viscera in order to provide 'mock' emotional and visceral feedback to contemplated actions. Though Damasio does not couch it in control theoretic terms, he is positing a visceral emulator, whose function is to provide mock emotional/visceral input – emotional imagery.

The present framework allows us to take Damasio's theory further than he takes it. If he is right that the brain employs a visceral/emotional emulator, then it is not only true that it can be used off-line, as he describes. It might also be used on-line as part of a KF scheme for emotional/perceptual processing. That is, just as perception of objects in the environment is hypothesized to involve an emulator-provided expectation that is corrected by sensation, so too emotional perception might involve expectations provided by the emulator and corrected by actual visceral input. And just as in environmental perception, the nature of the states perceived is typically much richer and more complex than, and hence underdetermined by, anything provided in mere sensation, so too the emotional emulator might be the seat of emotional learning and refinement, providing the ever maturing framework within which raw visceral reactions are interpreted.

Robert Gordon (Gordon, 1986) has been the primary champion of the 'simulation theory' in the 'theory of mind' debate in developmental psychology. The phenomenon concerns the development of children's ability to represent others as representing the world, and acting on the basis of their representations (Flavell, 1999; Wellman, 1990). The canonical example involves a puppet Maxi who hides a chocolate bar in location A, and then leaves. While out, another character moves the chocolate bar to location B. When Maxi returns, children are asked where Maxi will look for the bar. Children characteristically pass from a stage at which they answer that Maxi will look at B, to a stage where they realize that Maxi will look at A, since that is where maxi thinks it is. According to the simulation theory, we understand others' actions in this and similar situations by simulating them; roughly, putting ourselves in their situation and ascertaining what we would do. Such a simulation might well involve placing ourselves in another's perceptual situation (i.e. creating an emulated egocentric space/object situation), and perhaps their emotional situation with something like the emotion emulator discussed in the previous paragraph.

Lynn Stein (1994) developed a robot, MetaToto, that uses a spatial emulator in order to aid in navigation. The robot itself was a reactive system based on Brooks' subsumption architecture (Brooks, 1986, 1991). But in addition to merely moving around in this reactive way, MetaToto has the ability to engage its reactive apparatus with a spatial emulator of its environment in order to allow it to navigate more efficiently. By building up this map while exploring, MetaToto can then use this map both off-line (in a manner similar to Mel's models), and can also use it on-line to recognize its location, plan routes to previously visited landmarks, and so forth.

Applications to language are to be found primarily in the small but growing subfield of linguistics known as cognitive linguistics. The core idea is that linguistic competence is largely a matter of pairings of *form* and *meaning*; 'form' is typically understood to mean phonological entities, perhaps schematic, and 'meaning' is typically understood to be primarily a matter of the construction of representations similar to those enjoyed during perceptual engagement with an environment, especially objects, their spatial relations, force-dynamic properties, and perhaps social aspects as well. What sets this movement apart is a denial of any autonomous syntactic representation, and the notion that the semantics is based on the construction of representations more closely tied to perception than propositions.

Gilles Fauconnier (1985) has developed a theory of quantification, including scope and anaphoric phenomena, based on what he calls 'mental spaces', which at the very least analogous to spatial/object representations posited here. Ronald Langacker's Cognitive Grammar framework (1987, 1990, 1991, 1999) is a detailed examination of a breathtaking range of linguistic phenomena, including quantification (the account builds on Fauconnier's), nominal compounds, 'WH', passive constructions, and many dozens more. Karen van Houk (1995, 1997) has developed a very detailed account of pronominal anaphora within Langacker's Cognitive Grammar framework. Leonard Talmy (2000), and Lawrence Barsalou (Barsalou 1999; Barsalou et al., 1999) have also produced a good deal of important work in this area, all of it arguing forcefully that the semantic import of linguistic expressions consists in representations whose structure mimics, because derived from, representational structures whose first home is perception Exactly the sorts of representational structures made available by the various emulators described here.

5.4 Conclusion

The account I have outlined here is more schematic than I would ideally like. Ideally there would be both more detail at each stage, and there would be more evidence available in support of such details. In some cases such details and evidence have been omitted for reasons of space; in other cases the details and evidence are not currently extant. The primary goal, however, has been to introduce a framework capable of synthesizing a number of results and theories in the areas of motor control, imagery and perception, and perhaps even cognition and language. The synthesis shows how these processes are all interrelated as aspects of a single flexible information processing strategy.

In addition to this synthesizing potential, the model has another benefit. The fact that this strategy has its roots in motor control makes its phylogenetic appearance unmysterious, and thus renders the phylogenetic appearance of 'higher' representational functions based on it equally unmysterious.

These considerations are not theoretically insignificant, but they are also quite far from conclusive, or even, on their own, terribly persuasive. Ultimately, of course, informed and detailed investigation will determine the extent to which this framework has useful application in understanding brain function. To date, motor control is the only area in which

this framework has the status of a major or dominant theoretical player. I believe that part of the reason for this is that it is only in this area that theorists are generally familiar with the relevant notions from control theory and signal processing, and hence are thinking in terms of this framework at all.

Acknowledgements:

References:

Alain, C., Arnott, S. R., Hevenor, S., Graham, S., Grady, C. L. (2001). "What" and "where" in the human auditory system. *Proceedings of the National Academy of Science* USA 98(21):12301-6.

Barsalou, L. (1999). Perceptual Symbol Systems. *Behavioral and Brain Sciences* 22(4):577-609.

Barsalou, L., Solomon, K. O., & Wu, L. (1999) Perceptual simulation in conceptual tasks. In M. K. Hiraga, C. Sinha, & S. Wilcox (Eds.), *Cultural, Typological, and Psychological Perspectives in Cognitive Linguistics: The Proceedings of the 4ᵗʰ Conference of the International Cognitive Linguistics Association*, 3, Amsterdam: John Benjamins.

Behrmann, Marlene (2000). The mind's eye mapped onto the brain's matter. *Trends in Psychological Science* 9(2):50-54.

Blakemore, Sarah J., Susan J. Goodbody, and Daniel M. Wolpert (1998). Predicting the Consequences of Our Own Actions: The Role of Sensorimotor Context Estimation. *The Journal of Neuroscience* 18(18):7511–7518.

Brooks, R. A. (1986). A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, 2:14-23.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence* 47:139-160.

Chen, W., Kato, T., Zhu, X. H., Ogawa, S., Tank, D. W., Ugurbil, K. (1998). Human primary visual cortex and lateral geniculate nucleus activation during visual imagery. *Neuroreport* 9(16): 3669-74.

Craik, Kenneth (1943) *The Nature of Explanation*. Cambridge: Cambridge University Press.

Damasio, Antonio (1994). *Descartes' error : emotion, reason, and the human brain*. New York: G.P. Putnam.

Decety J, and Jeannerod M (1996). Mentally simulated movements in virtual reality. Does Fitts' law hold in motor imagery? *Behavioral Brain Research* 72:127-134.

Deiber, Marie-Pierre, Vicente Ibanez, Manabu Honda, Norihiro Sadato, Ramesh Raman, and Mark Hallett (1998). Cerebral Processes Related to Visuomotor Imagery and Generation of Simple Finger Movements Studied with Positron Emission Tomography. *Neuroimage* 7:73–85.

Desmurget, Michel, and Scott Grafton (2000). Forward modeling allows feedback control for fast reaching movements. Trends in Cognitive Sciences 4(11):423-431.

Farah MJ, Hammond KM, Levine DN, Calvanio R. (1988). Visual and spatial mental imagery: dissociable systems of representation. Cognitive Psychology 20(4):439-62.

Farah, M. J., M. J. Soso, et al. (1992). Visual angle of the mind's eye before and after unilateral occipital lobectomy. *Journal of Experimental Psychology: Human Perception and Performance* 18(1): 241-6.

Fauconnier, Gilles (1985). *Mental spaces: aspects of meaning construction in natural language*. Cambridge, MA: MIT Press.

Flavell, J.H. (1999). Cognitive development: children's knowledge about the mind. *Annual Review of Psychology* 50: 21-45.

Gelb, A. (1974). *Applied Optimal Estimation*. Cambridge, MA: MIT Press.

Gordon, Robert M. (1986). Folk psychology as simulation. *Mind and Language* 1:158-171.

Grush Rick (1995). *Emulation and Cognition*. PhD Dissertation, UC San Diego. UMI.

Grush, Rick (forthcoming). *The Machinery of Mindedness*.

Ito, Masao (1984). *The cerebellum and neural control*. New York: Raven Press.

Jeannerod Marc (1995). Mental imagery in the motor context. *Neuropsychologia* 33:1419-1432.

Jeannerod, Marc (2001). Neural Simulation of Action: A Unifying Mechanism for Motor Cognition. *NeuroImage* 14, S103–S109.

Jeannerod, Marc, and Victor Frak (1999). Mental imaging of motor activity in humans. *Current Opinion in Neurobiology* 9:735–739

Johnson, Scott H. (2000). Thinking ahead: the case for motor imagery in prospective judgements of prehension. *Cognition* 74 (2000) 33-70.

Kalman, R.E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(d):35-45.

Kalman, R., and Bucy, R.S. (1961) New results in linear filtering and prediction theory. *Journal of Basic Engineering* 83(d):95-108.

Kawato, Mitsuo (1999). Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology* 9:718–727.

Kosslyn, S. M., Alpert, N. M., Thompson, W. L., Maljkovic, V., Weise, S. B., Chabris, C. F., Hamilton, S. E., Rauch, S. L., & Buonanno, F. S. (1993). Visual-mental imagery activates topographically-organized visual cortex: PET investigations. *Journal of Cognitive Neuroscience*, 5, 263–287.

Kosslyn, S. M., Thompson, W. L., Kim, I. J., & Alpert, N. M. (1995). Topographical representations of mental images in primary visual cortex. *Nature*, 378, 496–498.

Krakauer, John W., Maria-Felice Ghilardi, and Claude Ghez (1999). Independent learning of internal models for kinematic and dynamic control of reaching. *Nature neuroscience* 2(11):1026-1031.

Lamm, Claus, Christian Windischberger, Ulrich Leodolter, Ewald Moser, and Herbert Bauer (2001). Evidence for Premotor Cortex Activity during Dynamic Visuospatial Imagery from Single-Trial Functional Magnetic Resonance Imaging and Event-Related Slow Cortical Potentials. *NeuroImage* 14, 268–283.

Langacker, Ronald W. (1987). *Foundations of Cognitive Grammar, Volume I.* Stanford: Stanford University Press.

Langacker, Ronald W. (1990). *Concept, image and symbol: the cognitive basis of grammar.* Berlin; New York: Mouton de Gruyter.

Langacker, Ronald W. (1990). *Foundations of Cognitive Grammar, Volume II.* Stanford: Stanford University Press.

Langacker, Ronald W. (1999). *Grammar and conceptualization.* Berlin; New York: Mouton de Gruyter.

Llinas, R., and Pare, D. (1991). On dreaming and wakefulness. *Neuroscience* 44:3:521-535.

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,* 562–571.

Mel, B. W. (1988). MURPHY: A robot that learns by doing. In *Neural information processing systems,* 544-553. New York: American Institute of Physics.

Pylyshyn, Z.W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition* 80(1-2):127-58.

Redon, Christine, Hay, Laurette, and Velay, Jean-Luc (1991) Proprioceptive control of goal-directed movements in man, studied by means of vibratory muscle tendon stimulation. *Journal of Motor Behavior* 23(2):101-108.

Richter, W., Somorjai, R., Summers, R. , Jarmasz, M., Menon, R. S., Gati, J. S., Georgopoulos, A. P., Tegeler, C., Ugurbil, K., Kim, S. G. (2000). Motor area activity during mental rotation studied by time-resolved single-trial fMRI. *Journal of Cognitive Neuroscience* 12(2): 310-20.

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

Schwartz, Daniel L. (1999). Physical Imagery: Kinematic versus Dynamic Models. *Cognitive Psychology* 38:433–464.

Stein, Lynn Andrea (1994). Imagination and situated cognition. *Journal of Experimental and Theoretical Artificial Intelligence* 6:393-407.

Talmy, Leonard (2000). *Toward a cognitive semantics* (2 volumes). Cambridge, MA: MIT Press.

Ungerleider, L. G., Haxby, J. V. (1994). 'What' and 'where' in the human brain. *Current Opinion in Neurobiology* 4(2): 157-65.

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.

van Hoek, Karen (1995). Conceptual reference points: A cognitive grammar account of pronominal anaphora constraints. *Language* 71(2):310-340.

van Hoek, Karen (1997). *Anaphora and conceptual structure*. Chicago: University of Chicago Press.

Van Pabst, J.V.L., & Krekel, P.F.C. (1993). Multi Sensor Data Fusion of Points, Line Segments and Surface Segments in 3D Space. *7th International Conference on Image Analysis and Processing*. (pp.174-182). Capitolo, Monopoli, Italy: World Scientific, Singapore.

Wellman, Henry M. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.

Wexler, M., Kosslyn, S.M., & Berthoz, A. (1998). Motor processes in mental rotation. *Cognition* 68:77-94.

Wolpert, D.M., Ghahramani, Z. and Jordan, M.I. (1995) An internal model for sensorimotor integration. *Science* 269:1880–1882.

Wolpert, Daniel M., Zoubin Ghahramani and J. Randall Flanagan (2001). Perspectives and problems in motor learning. *Trends in Cognitive Sciences* 5(11):487-494.

Yantis, S. (1992). Multielement visual tracking: attention and perceptual organization. *Cognitive Psychology* 24(3): 295-340.