BEHAVIORAL AND BRAIN SCIENCES (2004) **27,** 377–442
Printed in the United States of America

The emulation theory of representation: Motor control, imagery, and perception

Rick Grush
Department of Philosophy, University of California, San Diego, La Jolla, CA 92093-0119
rick@mind.ucsd.edu **http://mind.ucsd.edu**

Abstract: The *emulation theory of representation* is developed and explored as a framework that can revealingly synthesize a wide variety of representational functions of the brain. The framework is based on constructs from control theory (forward models) and signal processing (Kalman filters). The idea is that in addition to simply engaging with the body and environment, the brain constructs neural circuits that act as models of the body and environment. During overt sensorimotor engagement, these models are driven by efference copies in parallel with the body and environment, in order to provide expectations of the sensory feedback, and to enhance and process sensory information. These models can also be run off-line in order to produce imagery, estimate outcomes of different actions, and evaluate and develop motor plans. The framework is initially developed within the context of motor control, where it has been shown that inner models running in parallel with the body can reduce the effects of feedback delay problems. The same mechanisms can account for motor imagery as the off-line driving of the emulator via efference copies. The framework is extended to account for visual imagery as the off-line driving of an emulator of the motor-visual loop. I also show how such systems can provide for amodal spatial imagery. Perception, including visual perception, results from such models being used to form expectations of, and to interpret, sensory input. I close by briefly outlining other cognitive functions that might also be synthesized within this framework, including reasoning, theory of mind phenomena, and language.

Keywords: efference copies; emulation theory of representation; forward models; Kalman filters; motor control; motor imagery; perception; visual imagery

1. Introduction

The idea that one of the central tasks performed by the brain is to internally model various brain-external elements and processes is not new. In the twentieth century, Kenneth Craik was one of the most explicit proponents of this view (Craik 1943). In various guises the view has been taken up as an approach to understanding reasoning (Johnson-Laird 1983), theory of mind phenomena (Gordon 1986), mental imagery (Kosslyn 1994), and even aspects of motor control (Kawato 1999). The metaphor of "internal modeling" aside, these approaches have not (yet) been synthesized into anything approaching a unified and flexible framework.

In this article I have four goals. The first is to articulate an information-processing strategy that I will call the *emulation theory of representation.* This strategy is developed using tools from control theory and signal processing, especially drawing on pseudo-closed-loop control and Kalman filtering. I will try to use just enough mathematical formalism to ensure that the main ideas are clear. The second goal is to show the use of this framework in understanding certain aspects of motor control and motor imagery. Many researchers in the two fields of motor control and motor imagery currently appeal to constructs related to those I will develop, but such appeals rarely go into much detail concerning the overall information-processing structures involved, and little by way of clear synthesis has emerged. In providing such a structure, I hope to do a service to both of these areas of research by providing a framework within which various results can be synthesized, and within which a number of issues can be more clearly stated so as to avoid certain kinds of errors.

The emulation framework I will articulate not only allows for a great deal of motor control and motor imagery work to be synthesized in interesting ways, but it also synthesizes certain aspects of visual imagery and visual perception as well. Outlining such a synthesis is the third goal. The final goal, addressed in the last section, is to briefly explore the prospects for addressing other psychological capacities such as reasoning, theory of mind, and language, within the same framework.

Rick Grush is Associate Professor of Philosophy at the University of California, San Diego, the same university at which he received a joint doctorate in cognitive science and philosophy in 1995. His research in theoretical cognitive neuroscience is directed at understanding the nature of representation and the structure and neural implementation of fundamental representational capacities of nervous systems, including space, time, and objects. These topics will be explored in detail in his in-progress book, *The Machinery of Mindedness.*

© 2004 Cambridge University Press 0140-525X/04 $12.50
https://doi.org/10.1017/S0140525X04000093 Published online by Cambridge University Press

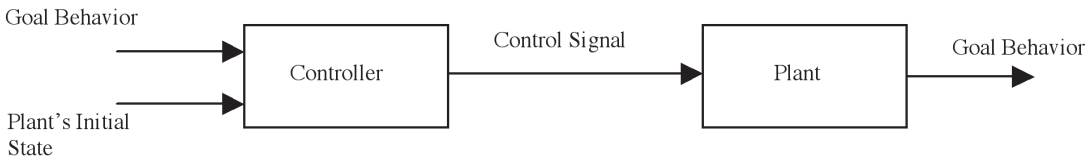


Figure 1. A simple feed-forward control schematic. "Plant" is the control theory term for the controlled system; in the case of motor control, the plant is the body, specifically the musculoskeletal system (MSS) and relevant proprioceptive/kinesthetic systems. Here, the "plant" includes the MSS, sensors,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

2. Motor control: Forward models and Kalman filters

2.1. Feed-forward and feedback control

The nature of the interaction between the motor centers of the brain and feedback from the body during fast, goal-directed movements has been a long-standing controversy in motor control (Desmurget & Grafton 2000; van der Meulen et al. 1990). 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. As a result of this motor volley, the body 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/kinesthetic feedback are used; the bulk of the motor sequence is determined and executed without feedback.

On the other side of the debate have been those who argue for *feedback* control. In the form most opposed to feed-forward control it is claimed that though there is a goal, there is no motor *plan* prior to movement onset. Rather, the motor centers continually compare the goal configuration to the current configuration (information about which is provided through vision or proprioception/kinesthesis), and simply move the current configuration so as to reduce the difference between it and the goal configuration. A simplified schematic for this sort of feedback control is shown in Figure 2.

In 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 causal 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.

Where the schemes differ is on *how* the controller implements the inverse mapping to produce the control sequence. In the feed-forward scheme, the entire motor plan is determined largely before movement onset. This often requires a good deal of sophistication on the part of the controller. In 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. (For more on these two schemes, especially the kinds of evidence historically used to argue in favor of each, see Desmurget & Grafton 2000.)

2.2. Emulators (forward models)

But feed-forward and feedback control do not exhaust the alternatives. There has been growing recognition among researchers in motor control that schemes involving the use of *forward models* of the musculoskeletal system (henceforth MSS) are promising (Kawato 1999; Wolpert et al. 2001). I will use "emulator" as a more descriptive synonym

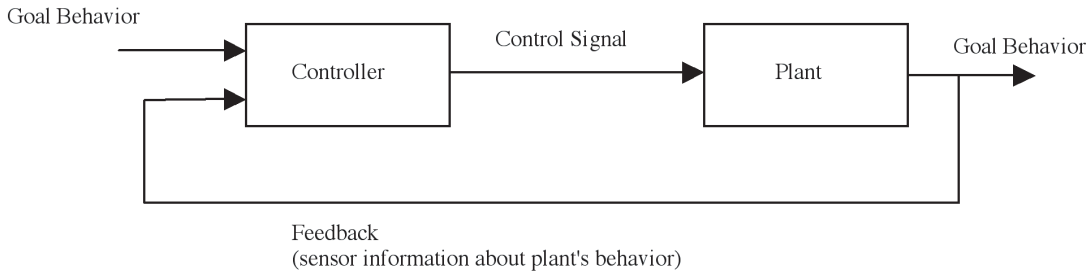


Figure 2. A simple feedback control schematic. Sensors (a component of the plant) measure critical parameters of the plant, and this information is continually provided to the controller. The controller uses this feedback in its evolving production of the control signal.

https://doi.org/10.1017/S0140525X04000093 Published online by Cambridge University Press

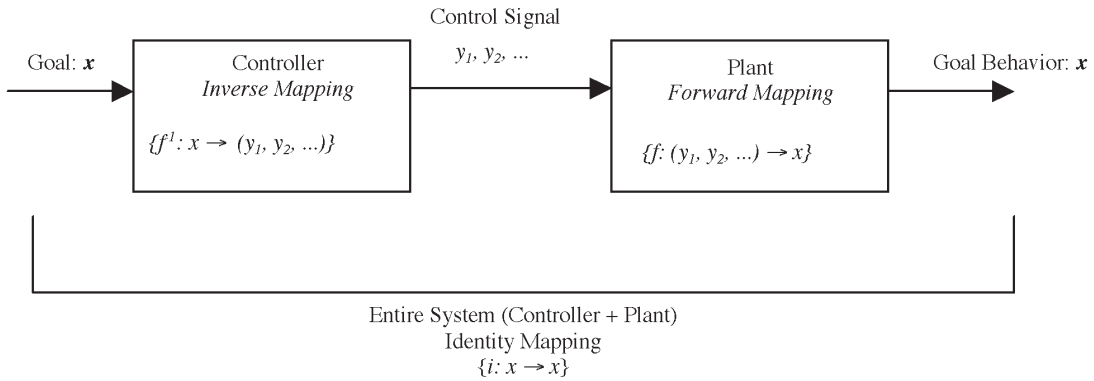


Figure 3. Forward and inverse mappings.

for "forward model." In such schemes, the controller exploits a continuing stream of feedback provided by an emulator of MSS dynamics, which is driven by efference copies of the motor commands sent to the body. The simplest such scheme is shown in Figure 4.

In this scheme the controlling 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. Such feedback has a number of uses. One use is that emulator feedback can be subject to less delay than feedback from the periphery (Desmurget & Grafton 2000; Wolpert et al. 1995). This will be explored in more detail in section 2.5.

There are two points to highlight that will be important later on. First, for these specific purposes it does not matter *how* the emulator manages to implement the forward mapping (though of course for other purposes it matters quite a bit, and empirical evidence is relevant to deciding between the options I will discuss below). One possibility is that an emulator 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 (see Miles & Rogers 1993). Another possibility is for the emulator to be what I will call an *articulated model*. The *real* MSS behaves the way it 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, for example, stretch receptors and Golgi tendon organs. This measurement constitutes bodily proprioception and kinesthesis: 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 inertia, such that, just as an

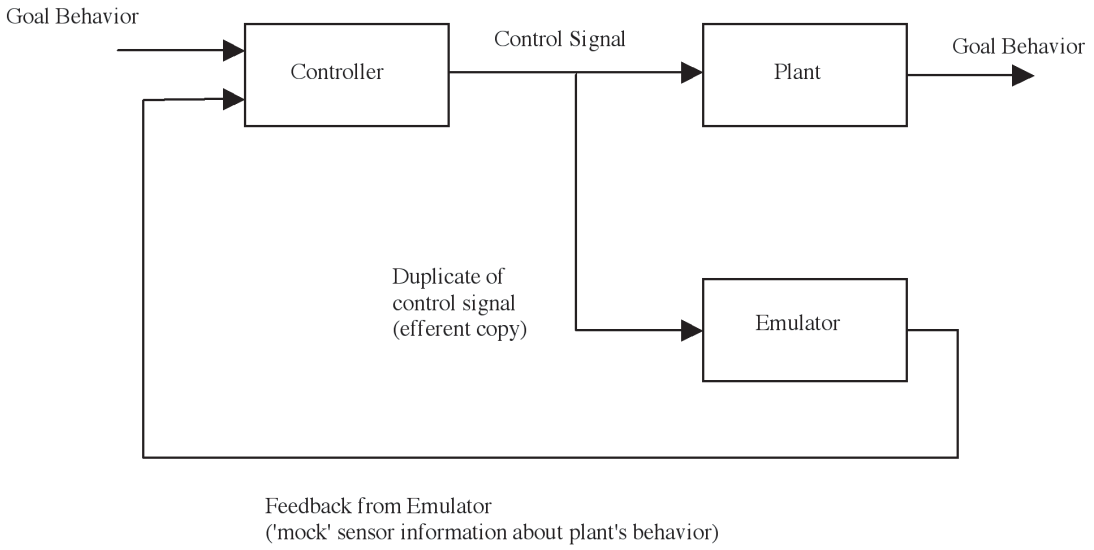


Figure 4. A simple pseudo-closed-loop control schematic (Ito 1970; 1984). A copy of the control signal – an *efference copy* – is sent to a subsystem – an emulator – that mimics the input-output operation of the plant. Because the emulator is given the same input as the plant (its input is a copy of the signal sent to the plant) it produces a similar output.

https://doi.org/10.1017/S0140525X04000093 Published online by Cambridge University Press

increase in elbow angle results in an increase in arm angular inertia, an increase in the firing rate of the first group of neurons instigates an increase in the firing rate of the second (see Grush 1995; Kawato et al. 1987). And just as the real MSS is subject to a measurement that provides proprioceptive and kinesthetic information, the articulated emulator can have a "measurement" taken of the same variables, and thus yield a mock sensory signal. There are more than just these two options, but for purposes of this article something more along the lines of the articulated emulator will be assumed – though this is an empirical issue, of course.

The second point is that emulators, whether lookup tables or articulated models, must, under normal conditions, have a certain degree of plasticity. This is because the systems they emulate often alter their input-output function over time. This phenomenon is *plant drift* – in the case of mechanical plants, belts loosen, gears wear, some parts get replaced by others that are not exactly the same; in the case of the body, limbs grow, muscles get stronger or weaker over time. Whatever the details, a particular motor command might lead to one MSS 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.

2.3. 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 as it stands to be of much use in explaining real biological systems. The next conceptual ingredient needed in order to bring these basic ideas up to speed is the Kalman filter (which I will abbreviate KF). 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 to introduce those aspects of KFs that are important for the remaining discussion.[1] The technique, a standard version of it anyway, is diagrammed in Figure 5.

First, we need a description of the problem to be solved by the KF. The problem is represented by the top part of the diagram, within the dashed-line box. We start with a *process* consisting of a system of k state variables, and whose state at time t can thus be described by the $k \times 1$ vector $r(t)$. The process' state evolves over time under three influences: first, the process' 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 or the driving force, the process' state at any given time would be a function of its state at the previous time: $r(t) = Vr(t\text{-}1)$, where V is a $k \times k$ 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 $k \times 1$ vector $n(t)$, and the driving force as another $k \times 1$ vector $e(t)$. Thus, the process at time t is: $r(t) = Vr(t\text{-}1) + n(t) + e(t)$.

The real signal $I(t)$ is a *measurement* of states of this process. We can represent this measurement as a $k \times h$ measurement matrix O, that maps $r(t)$ to the $h \times 1$ signal vector $I(t)$: $I(t) = Or(t)$. (An obvious special case is where O is the identity matrix I, in which case $I(t) = r(t)$.) We can

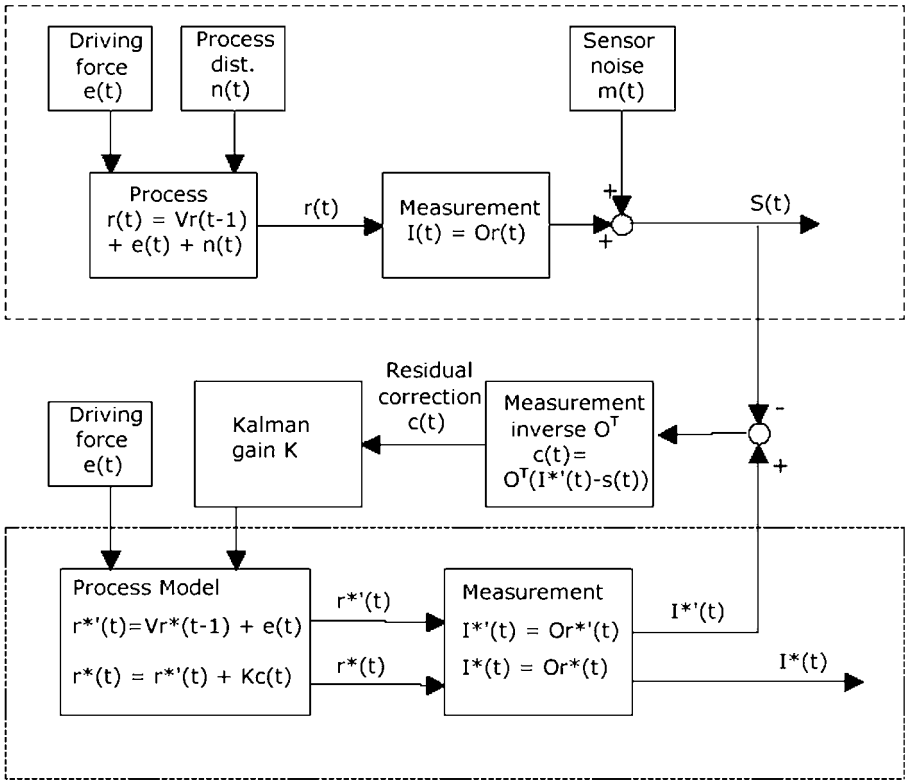


Figure 5. A basic Kalman filtering scheme.

https://doi.org/10.1017/S0140525X04000093 Published online by Cambridge University Press

represent unpredictable noise in the measurement process – the sensor noise – as another small time-dependent zero-mean $h \times 1$ 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.

The core idea behind the KF is that it maintains an optimal estimate of the real process' state, and then subjects this estimate to the same measurement that produces the real signal from the real process. The result is an optimal estimate of the noise-free signal. The KF's estimate of the process' state is embodied in the state of a *process model*, an articulated model of the process. 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, the value of each of the k state variables of r^* to the corresponding state variable of r.

So far it should be clear how, with the benefit of an accurate state estimate $r^*(t)$, the KF can produce an optimal estimate of the real signal $I(t)$. Now we must examine how it maintains an optimal state estimate $r^*(t)$. This is done in two steps. The first step is often called the *time update*. Given the previous state estimate $r^*(t\text{-}1)$, the KF produces an expectation or prediction of what the state at t will be by evolving $r^*(t\text{-}1)$ according to the same dynamic V that governs the evolution of the real process and adding the driving force $e(t)$.[2] The result of this time update is the *a priori* estimate, $r^{*\prime}(t)$ (note the prime), and as stated, it is arrived at thus: $r^{*\prime}(t) = Vr^*(t\text{-}1) + e(t)$. It is called the *a priori* estimate because it is the estimate arrived at before taking information from 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, and also given my knowledge of the driving force, what should I expect the next state to be?" This is the a *priori* estimate.

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 $r^{*\prime}(t)$ is measured to produce an a priori signal estimate $I^{*\prime}(t)$. 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. This is done by pushing the residual through the inverse of the measurement matrix O, O^{T}. This is the *residual correction* $c(t)$. Though the KF now knows exactly how to alter the a priori estimate in order to eliminate the sensory residual, it typically does not apply the entire residual correction. Why? The residual correction is how much the a priori state estimate would have to be altered to eliminate the sensory residual. But not all of the sensory residual is a result of the inaccuracy of the a priori estimate. Some of it is a result of sensor noise. Thus, the a priori estimates $r^{*\prime}(t)$ and $I^{*\prime}(t)$ might be very accurate, and the sensory residual due mostly to the sensor noise.

The KF determines how much of this correction should actually be applied 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 than the observed signal, and so a smaller portion of the residual 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 $I(t)$.

2.4. 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 control signal. Everything in the dashed-line box is functionally equivalent to the plant in the pseudo-closed-loop scheme of Figure 4, and everything in the dotted-line box is functionally equivalent to the emulator. The box labeled "plant" in Figures 1–4 did not separate out the process, sensors (a.k.a. measurement), and noise, but lumped them all together. The "emulator" box in Figure 4 similarly lumped together the emulation of the process and the emulation of the measurement via sensors.

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 degenerate cases. If the Kalman gain is set so that the entire sensory residual is always 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 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. The state estimate of the process model (emulator) evolves over time exclusively under the influence of its own inner dynamic and the controller's efference copies, and the feedback sent to the controller is just a current measurement of this encapsulated estimate, just as in pseudo-closed-loop control.

I will refer to systems like that in Figure 6 as KF-control schemes. At the same time, I will indulge in a certain degree of flexibility in that I will take it that extensions to con-

https://doi.org/10.1017/S0140525X04000093 Published online by Cambridge University Press

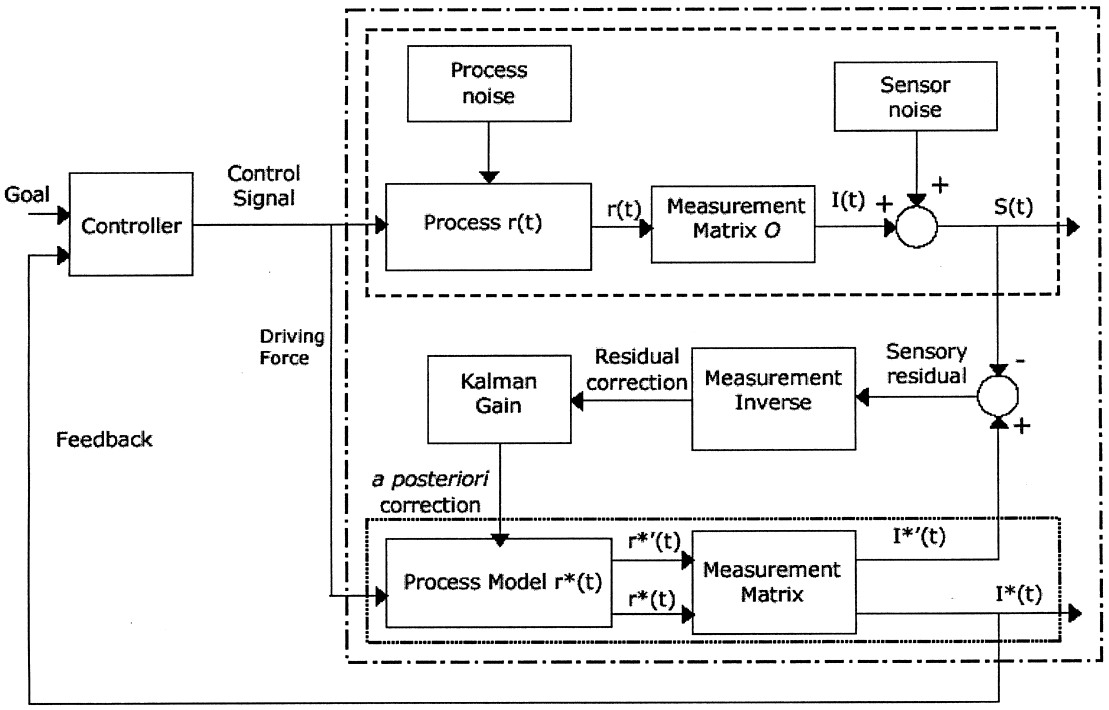


Figure 6. A control scheme blending pseudo-closed-loop control and a Kalman filter.

tinuous and nonlinear systems are included (see the references in Note 1 for more on these cases), 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 sect. 3).

Before closing this section, it might help to provide a qualitative example of a problem structure and solution structure that take the form described by the formalism above. The example is ship navigation.[3] The "process" is the ship and its location, which at any time is a function of its location at the previous time, plus the driving force (the engine speed, rudder angle), plus the "process noise" (unpredictable winds and ocean currents). The "signal" in this case is bearing measurements, which consist *inter alia* of the observed angles to known landmarks or stars. The observation-measurement process is not perfect, and this imperfection can be effectively represented as "sensor noise" added to an ideal bearing measurement. The task – exactly the same as the Kalman filter's task as described in section 2.3 – is to determine, on the basis of the imprecise bearing measurements, the ship's actual position (or equivalently, what the noise-free, or perfect, bearing measurements would be). To solve this, the navigation team maintains a model of the process, implemented on a map. At each time step (or "fix cycle"), the team projects where the ship should be, on the basis of the estimate from the previous fix cycle and the known driving force. This a priori prediction is then combined with the actual observations in order to determine a "best guess" region within which the ship is located: the a posteriori estimate. And this a posteriori estimate becomes the initial estimate for the next cycle. And, just as in the KF, if an ideal navigator knows that there are no unpredictable winds or currents, but the bearing measurements are very imprecise, then the prediction based on the previous location – the a priori estimate – will be weighted more heavily than the bearing measurements (low Kalman gain). If, on the other hand, the bearing measurements are quite reliable but there are a lot of winds and currents (high process noise), then the "sensory residual" will be weighted more heavily (high Kalman gain). Of course, real navigators do not calculate anything like a Kalman gain matrix for optimally combining a priori expectation and observed signals. Real navigators are often good, but seldom optimal. This is one of the many ways in which the emulation framework relaxes the strict requirements of the Kalman filter, while maintaining the general information-processing structure.

The general structure of the problem and the general structure of the KF's solution – a structure captured by the emulation framework – are quite unmysterious. One system (a ship's crew, a brain) is interacting with another system (a ship, an environment, a body) such that the general principles of how this system functions are known, but the system's state is not *entirely* predictable owing to process noise (unpredictable currents, bodily or environmental perturbations), and imperfect sensors. The solution is to maintain a model of the process – done by a part of the ship's crew, the navigation team; or specialized emulator circuits in the brain – in order to provide predictions about what its state will be; and to use this prediction in combination with sensor information in order to maintain a good estimate of the actual state of the system that is being interacted with – an estimate that makes use of sensor information, but is not limited by the many sub-optimalities of the bare sensory process.

2.5. Evidence for the emulation framework

The emulation framework is a specification of a certain kind of information-processing structure. It is my hypothesis that the brain uses this information-processing structure for a number of tasks, and it is also my contention that this hypothesis, if true, can serve to synthesize a wide range of oth-

https://doi.org/10.1017/S0140525X04000093 Published online by Cambridge University Press

erwise seemingly unrelated models and results. My evidence for the claim that the brain uses the emulation framework is distributed throughout this paper, but in most cases it takes the following form. First, I introduce some specific area of potential application, such as motor control or visual imagery. I then pick out some exemplar model or result, already existing in the cognitive neuroscientific literature in this particular domain, which offers a particularly good example of how the emulation framework applies in that domain. Because the exemplar model is a specific application of the broader synthesizing emulation framework, the evidence, typically produced by the researchers who proposed the specific model, becomes evidence for the emulation framework's application to that domain. And vice versa, any evidence that would count against the specific exemplar model in question would, prima facie, count as evidence against the emulation framework's applicability to that domain. Since it is the goal of this article to articulate the synthesizing framework, which requires articulating the framework and applying it to a number of domains, space limitations prevent me from going into anything approaching a convincing level of detail concerning the evidence for the model in each of the areas covered. Rather, my goal is to cover just enough of the evidence, typically in the form of one or two exemplar models or experimental results, to make the character of the proposed synthesis clear – to give specific flesh to not only how the emulation framework applies to that domain, but also to indicate the kinds of evidence that count for and against the emulation framework in that domain. For those who wish to see more detailed evidence within each domain of application, I will provide references to specific studies and models, done by specialists within the specific field.

2.6. Motor control

The first application is within the domain of motor control, and the exemplar of an application of the emulation framework to motor control I use here is a model by Wolpert, Ghahramani, and Jordan (Wolpert et al. 1995). The goal of their project was to provide a model that would explain the information-processing mechanisms used by the nervous system to maintain an estimate of the body's state during movements. Wolpert et al. distinguished three possibilities: pure peripheral sensory information, pure centrally generated predictions based on motor outflow (a.k.a. knowledge of the driving force via efference copies), and a combination of these two integrated via a Kalman filter scheme essentially identical to that described in section 2.4. The data were subjects' estimates of the position of their own hands after movements of varying lengths of time executed without visual feedback. The observed pattern demonstrated that as the duration of the movement increased from .5 second (the shortest movement duration) to 1 second, the magnitude of subjects' over-estimation of the position of their hands increased. But after a maximum at about one second, the magnitude of the overestimation decreased. This same temporal pattern was observed in three conditions: assistive force, resistive force, and no external force.

Wolpert et al. argued that neither the pure sensory inflow nor the pure motor outflow-based models could account for this pattern, but that the Kalman filter model could. Because proprioceptive and kinesthetic feedback is not available during the initial stages of the movement, the state estimate early on is based almost entirely on the uncorrected predictions of the process model. Initial error (for which there is an initial overestimate bias) in this estimate compounds as time progresses in absence of correction from sensory signals. However, as feedback becomes available from the periphery, this information is used to make corrections to the state estimate that begin to show their influence about one second after movement onset. Hence, there is a drop in the magnitude of the overestimation after about one second.

Many models and experimental results published in recent years provide strong evidence in support of the idea that at least some aspects of motor control are subserved by some kind of KF information-processing structure. These include Blakemore et al. (1998), Kawato (1999), Wolpert et al. (2001), Krakauer et al. (1999), and Houk et al. (1990). Mehta and Schaal (2002) study the dynamic characteristics of subjects' movements in the pole balancing task (including "blackout" trials where visual feedback is blocked for some brief period), and show that the data strongly suggest that a Kalman filter-like mechanism is used by the central nervous system to process sensory information during the execution of this task. As for what evidence would count against the emulation framework's applicability to motor control, it is whatever evidence would count against the models just cited. For example, a lack of a time-varying estimation pattern as found by Wolpert et al. (1995), or behavioral patterns of pole balancing tasks as described in Mehta and Schaal (2002), which would disconfirm a Kalman filter mechanism and implicate instead any of the other possibilities they claim are disconfirmed.

3. Motor imagery

3.1. 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 errors. This is part of what allows the KF to be an optimal state estimator, and in control contexts having optimal information is often 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 sensory/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 final percept is determined by the expectation, which in turn is a function of the operation of the model's inner dynamic as driven by efference copies, if any. The higher the Kalman gain, the more this inner dynamic is overridden by sensory deliverances. That is, the same system not only implements both sorts of processing, but can flexibly and optimally combine them as conditions and context dictate.

Section 5 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 so that real sensory information

https://doi.org/10.1017/S0140525X04000093 Published online by Cambridge University Press

has *no* effect. The emulator's state is allowed to evolve according to its own dynamic and the efference copies, if any; there is no "correction" or alteration from the senses. Second, the motor command must be suppressed from operating on the body.[4] Therefore, 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 musculature, and with the normal sensory inflow having no effect on the emulator's state and feedback. The very same process that is used to enhance overt motor control, as described in section 2.6, can thus be used to produce motor imagery. In this section I will briefly defend this emulation theory of motor imagery. The purpose is not only to give another example of the framework in operation, but also to show how it can help to clarify some of the issues within the field.

It will be helpful to compare the emulation theory of motor imagery to another proposal that superficially seems identical, but which is actually quite distinct. I have in mind the "simulation theory" of motor imagery, currently favored by a number of researchers (Jeannerod 2001; Johnson 2000a), according to which motor imagery is just the inner off-line operation of the efference motor centers of the brain. As Jeannerod and Frak put it, "motor imagery corresponds to a subliminal activation of the motor system" (Jeannerod & Frak 1999).

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 – efferent 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 motor imagery hypothesized to be the free-spinning of the controller (motor centers) when disengaged from the plant (body). In the emulation theory, by contrast, imagery is not produced by the mere free-spinning operation of the efferent motor areas, but by the efferent motor areas driving an emulator of the musculoskeletal system. The next section quickly recaps two areas of evidence typically cited in support of the simulation theory. In section 3.3, I will discuss considerations favoring the emulation theory over the simulation theory. Before I move on, I should clarify something. I will use the terms "simulation theory" and "emulation theory" exactly as I have defined them in this paragraph. Given these definitions, it should be clear that the simulation theory and the emulation theory are not at all the same thing. They agree that the efferent motor centers are active during imagery. The simulation theory takes this by itself to be sufficient for motor imagery; the emulation theory does not, and claims that in addition, an emulator of the musculoskeletal system is needed and imagery is produced when the efferent motor centers drive this emulator. This distinction should be entirely obvious. To make an analogy: The emulation theory claims that motor imagery is like a pilot sitting in a flight simulator, and the pilot's efferent commands (hand and foot movements, etc.) are translated into faux "sensory" information (instrument readings, mock visual display) by the flight simulator which is essentially an emulator of an aircraft. The simulation theory claims that just a pilot, moving her hands and feet around but driving neither a real aircraft nor a flight simulator, is sufficient for mock sensory information.

It may be the case that some of the researchers I cite as defending the simulation theory really mean something like the emulation theory, but just are not entirely clear about the distinction in their descriptions. If that is the case, then maybe even the defenders of the "simulation theory" are really agreeing with the emulation theory. In what follows I will compare the simulation theory as I describe it with the emulation theory as I describe it. If nobody really holds the simulation theory (as I define it), then at least some clarity will have been added to the literature on this topic.

3.2. *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 are active during motor imagery, the second concerns a number of isomorphisms between overt movements and their imagined counterparts.

That motor imagery involves the off-line activity of many motor areas is a widely replicated result (for a recent review, see Jeannerod 2001). PET studies of motor imagery consistently show selectively increased activity in premotor areas, supplementary motor areas, the cerebellum, and other areas. This is the defining feature of the simulation theory: that motor imagery is the psychological counterpart of the off-line operation of the efferent motor areas. Of the major motor areas canvassed in such studies, only primary motor cortex is (usually) conspicuously silent during motor imagery. This would seem to imply that the major signal bifurcation, where the efference copy is split from the "real" efferent signal, occurs just before primary motor cortex. (Though the implementation details are not entirely clear: Some studies, such as Richter et al. [2000], implicate primary motor cortex in motor imagery. Therefore, the functional separation of the efferent motor areas and the MSS emulator may not be easy to determine.)

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 (Deiber et al. 1998; Jeannerod & Frak 1999). 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 (Jeannerod 1995). The frequency at which subjects are no longer able to overtly move their finger between two targets is the same for overt and imagined movement. There is evidence that even Fitts' Law is in effect in the domain of imagination (Decety & Jeannerod 1996).

Johnson (2000a) has provided compelling evidence that, when forming a motor plan, a subject's expectations about what range of grip orientations will be comfortable is very close to the range that actually is comfortable. Johnson argues that this result indicates 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), but also that such imagery, precisely because it respects biomechanical constraints, is used to determine an effective motor plan before execution.[5]

The proponents of the simulation theory point out that

https://doi.org/10.1017/S0140525X04000093 Published online by Cambridge University Press

not only are motor areas active during motor imagery, but the isomorphisms 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.

3.3. Emulation versus simulation

None of the results appealed to in the previous section distinguish the simulation from the emulation theories, as both expect the controller (efferent motor centers) to be active during the production of motor imagery. The difference is that the emulation theory claims that mere operation of the motor centers is not enough; to produce imagery, the motor areas must be driving an emulator of the body (the MSS and relevant sensors).

There are reasons for preferring the emulation theory – for thinking that the mere operation of the motor centers is insufficient for motor imagery. 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 and kinesthesis. *These two things are not the same.* Oddly, this point seems to be underappreciated, so it is worth stressing. In the case of overt movement, motor volleys are sent to the musculature, and this results in movements of the limbs that can be specified dynamically or kinematically. As a functionally distinct part of the movement process, various kinds of sensors, such as stretch receptors and Golgi tendon organs, have the job of measuring aspects of the body in motion. Such measurements result in proprioception and kinesthesis. There is no reason to think that the result of measurement by these sensors produces a signal with even the same number of parameters as the dynamic or kinematic signal, let alone that they are in the same format (i.e., a format such that one can be used as a substitute for the other). The only way to get from the former (signals in motor format) to the latter (signals in proprioceptive and kinesthetic format) is to run the motor signal through something that maps motor plans/signals to proprioception and kinesthesis. And the two possibilities are (a) the body (yielding real proprioception and kinesthesis), and (b) a body emulator (yielding faux proprioception and kinesthesis).

That a motor plan and a sequence of proprioceptive/kinesthetic feelings are distinct should be obvious enough, but the difference can be brought out rather nicely by comparing how the simulation theory and the emulation theory account for phantom limb phenomena. 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 (Vilayanur Ramachandran, personal communication). Recall the point made in section 2.2 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 had a long period where it is being told that the correct input-output mapping is a degenerate many-to-one mapping that produces the same unchanging proprioceptive/kinesthetic signal 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.[6]

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/kinesthetic sensations produced when it is executed, is another. The simulation theorist must maintain that those who have an inability to produce motor imagery of a certain sort (because their phantom limb is paralyzed) also have an inability to produce motor plans of that sort. But subjects with paralyzed phantoms seem able to make motor plans – they know all too well what it is that they cannot do; they cannot move the limb from their side to straight in front, for example. What is wrong is that when the plan is executed on the phantom, nothing happens: the proprioceptive/kinesthetic signal remains stubbornly static.[7] A motor plan is one thing, a sequence of proprioception and kinesthesis is another. The simulation theorist conflates them. This issue will be revisited in the case of visual imagery, where the difference between a motor plan and the sensations resulting from the execution of that plan is clearer.

3.4. Neural substrates

The question of where the neural substrates of the MSS emulators are located is natural. (These emulators are used both in motor control and motor imagery, hence my discussion in this subsection will make reference to both cases.) Masao Ito, who first proposed that emulators would be involved in motor control, postulated that they were in the cerebellum, based on behavioral and lesion data (Ito 1970; 1984). Daniel Wolpert and others (Wolpert et al. 2001) review a range of neurophysiological evidence that also implicates the cerebellum. Motor imagery has long implicated the cerebellum as well as other motor areas. Many such studies are cited in Imamizu et al. (2000), Jeannerod (1995; 2001), and Jeannerod and Frak (1999). In a recent paper that contains useful citations to further neurobiological investigations of MSS emulators, Naito et al. (2002) present results from a range of neuroimaging studies of motor imagery they have conducted, showing that the cerebellum (along with other areas) is selectively active during motor imagery.

4. Visual imagery

So far the applications of the strategy of emulation have been only with regard to motor control and motor imagery. But the same basic scheme is obviously applicable to imagery from other modalities, provided that they have some relevantly exploitable structure. Section 4.1 introduces two models by Bartlett Mel (Mel 1986; 1988), robots that can

https://doi.org/10.1017/S0140525X04000093 Published online by Cambridge University Press

use imagery to solve certain problems. Though Mel does not describe them in this way at all, they generate imagery by operating emulators of the motor-visual loop. The details of how these systems generate imagery are 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. After getting the basic idea across with Mel's models in section 4.1, I will turn in sections 4.2 and 4.3 to providing evidence that some aspects of visual imagery are generated via the operation of a motor-visual emulator in a way at least roughly analogous to that suggested by Mel's models. In section 4.4., 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 arrangements of solid objects and surfaces in egocentric space. I show how the two forms of emulation can work in tandem. My goal in these sections is not to provide anything like a systematic or complete theory of visual imagery, but rather to indicate in broad outline how some of the more prominent theories of visual imagery can be seen as implementations of the emulation theory.

4.1. 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 that 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 represented. Murphy operates in two modes. In the first mode, it 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 on 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 is that each unit in the visual grid is actually a connectionist unit that receives an input not only from the video camera, as described, but also receives connections from neighboring units, and a copy of Murphy's motor command (e.g., increase elbow angle, decrease shoulder angle). During overt movement, the units then learn to associate these patterns of input with the future inputs they receive from the video camera – they 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) to see what outputs the system produces on their basis (in this case, the 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, and the use of this ability is the second mode of Murphy's operation. 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 arm and camera off-line, and manipulates the visual grid with efference copies. It moves the image of its arm around by means of the same motor commands that would usually move its arm around, seeing what sequences of movement impact upon 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 and implements, in one go, the solution.

Mel nowhere puts any of this in terms of control theory or forward mappings, et cetera. Rather, he describes it simply as a connectionist network that learns to solve problems through imagery. Nevertheless, during the imagery phase it is 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 real 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 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. In 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 intelligent anticipation would be crucial in interpreting the input.

The next model, also by Mel, is similar to Murphy.[8] 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 also from efference copies and from connections to other units in both grids. Again, the grids learn the forward mapping of the motor-visual loop by learning to associate current visual inputs plus motor commands with future visual inputs. 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 efference 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 that would (overtly) move towards or away from the object.

https://doi.org/10.1017/S0140525X04000093 Published online by Cambridge University Press

As with Murphy, Mel does not couch any of this in terms of control loops or emulators. And again, the potential for exploiting the grids as part of a Kalman filter-like mechanism for processing perception is not explored.[9]

4.2. Visual imagery and visual areas of the brain

In 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 are 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 are active, but that their activity is selectively similar to the activity of such areas during the analogous overt perceptual situations. Because the focus is currently on imagery that is modality specific (see sect. 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; see also Kosslyn et al. 1993). Kosslyn et al. (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 it presents are 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 as in the case of Mel's models, the image is actually produced on a topographically organized medium, and manipulated via efference 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.

4.3. 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 active during, and crucial to, certain sorts of visual imagery. Activity in premotor areas has been widely shown to occur during visual imagery tasks requiring image rotation. Richter et al. (2000) demonstrated this with time-resolved fMRI, a result confirmed by Lamm et al. (2001).

Though 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.

In these experiments, subjects had to solve problems already known to involve certain kinds of visual imagery, specifically the mental 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 on the imagery task was much better than in trials in which the directions were different. On the emulation theory of imagery, the result is expected. Part of what would be required in order to emulate an event of rotating a shape is a covert motor command such as twisting with one's hand. This is presumably more difficult to do when the motor systems are actually producing a different command than when they are actually producing the required command.[10]

A check written at the very end of section 3.3 can now be cashed. As mentioned there, a motor command is one thing, and the character of the sensory signal it produces is something else. Even if this were not clear in the case of motor imagery, it should be quite clear in the case of visual imagery. While motor areas are involved in imagery as described above, clearly the motor plan *by itself* cannot determine 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." Exactly this underdetermination is present in the case of motor imagery, though it is more difficult to recognize here because the body is relatively stable (which is why cases involving plant drift, paralysis of phantom limbs, etc., are good ways to bring the point out).

4.4. Modal imagery versus 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 concern 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 have taken a different form.

https://doi.org/10.1017/S0140525X04000093 Published online by Cambridge University Press

For example, it might have taken 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 overall 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 is relatively stable, and hence can be learned and 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 learned and 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 – the topographic visual input grid or the arm and workspace – 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 different ways. The Mel type systems produce it by constructing an emulator of the sensory input grid itself, and they do this by letting the same hardware act as both the emulator and the sensory topographic input grid. In such a case, the emulator's state is just the visual image, there is no measurement; or if you like, the measurement consists of the identity matrix. The other potential system I described produces visual imagery by constructing and maintaining an emulator of the *arm's* state and the *workspace'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 the dashed-line box, 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, in the box with the dot-and-dash line, corresponds to a modality-specific emulator, as exemplified in Mel's models. This emulator's states are just 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. Given that the emulated system is just the visual input medium, no measurement is needed in order to arrive at a visual image.

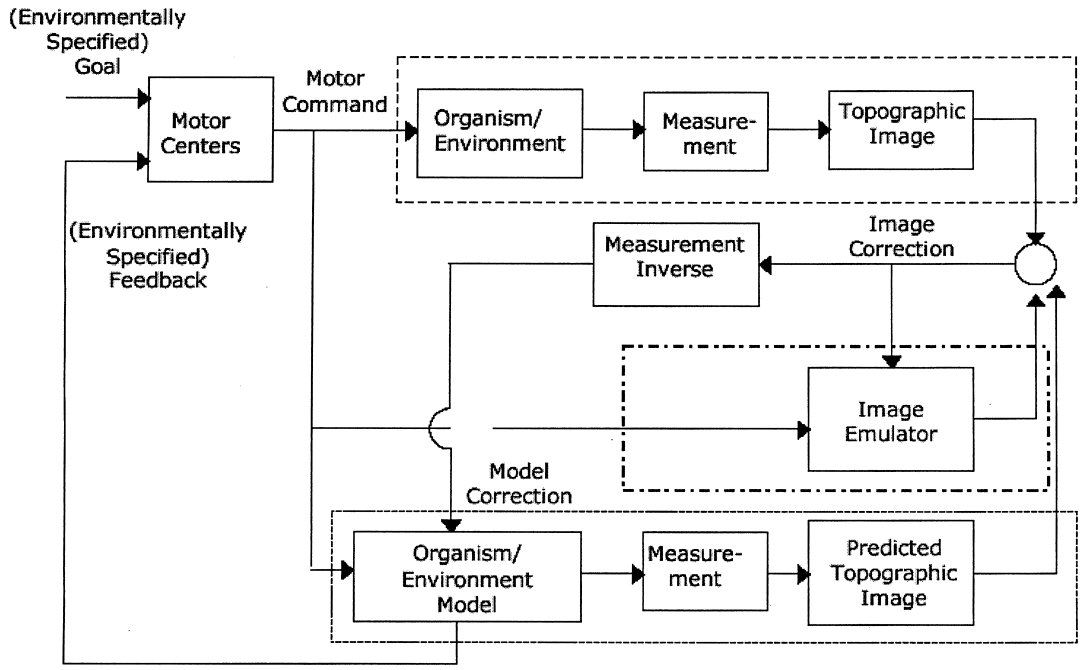


Figure 7. A KF-control scheme using two emulators: one a modality-specific "image" emulator; the other an amodal organism/environment emulator.

https://doi.org/10.1017/S0140525X04000093 Published online by Cambridge University Press

The third, in the dotted-line box 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 sect. 5).

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 top dashed-line box by pointing out that a white cube is two meters in front of the organism, which is causing a square pattern of stimulation on the retina. The creature moves to the right, and so the cube is now slightly to the creature's left (this is the change in the "organism/environment" box), 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" efference 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. Given that the emulator in this case 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 bottom dotted-line box 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 a square topographic image. The "move right" efference 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 (this is the change in the "organism/environment model"). Subjecting this model to a "visual measurement" yields a new topographic input image similar to the previous one, only with the patterns altered slightly. When operated off-line, the result would be visual imagery, if measured; or amodal imagery, if not subjected to a measurement. When 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 (completely) tied to the contingencies of 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 modality-specific 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 will not bother with a diagram, which would be rather cumbersome, 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 (Alain et al. 2001; van Pabst & Krekel 1993).

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 behavioral environment without a corresponding modality-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 5 and 6.[11]

4.5. 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 do not obviously involve any sort of motor component, as when one simply imagines a vase of flowers.

An emulator by itself does not decide when it gets operated on-line versus 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 emulator system. Of course other processes are required. But they would be required in 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

https://doi.org/10.1017/S0140525X04000093 Published online by Cambridge University Press

outside the scope of the present focus. The initial state of the emulator gets seeded somehow, perhaps with a memory of a state it was once in. An emulation 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 (see next section), it is using emulators as part of a KF-control scheme. The fact that there are connections to broader cognitive components is 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.

In 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 efference copies of motor commands. In this section I have focused on imagery produced through the emulation of processes that have no (or a minimal) dynamic of their own, but depend for their evolution on being driven by efference copies. Mel's models highlight this, as does the imagery involved in Wexler's studies. But this has been merely 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 efference 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 efference copies, but by the modeling of processes that evolve over time on their own (Kosslyn [1994] refers to this sort of dynamism in imagery as "motion encoded," as opposed to "motion added," which is driven by efference copies). The model thus includes these other sorts of imagery as degenerate cases. I focus on the case involving efference 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 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, perhaps exploiting early visual areas. 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, perhaps early visual areas, as well as systems in the dorsal stream. And some tasks requiring spatial imagery might not involve any notably visual component at all.[12] 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, because, surprisingly, the dominant accounts of motor and visual imagery in the literature are not at all parallel. Dominant views on motor imagery, as we have seen, equate it with the covert operation of *efferent* processes – either an efference copy or a motor plan. Dominant views on visual imagery treat it as the covert "top-down" stimulation of *afferent* areas. That these two dominant explanations seem so prima facie different is at best surprising, at worst embarrassing. First, intuitively, both kinds of imagery seem to be cases of a similar kind of process, and this is no doubt why they are both called "imagery." Second, a unified account is more theoretically parsimonious, showing how different phenomena can be explained by a single framework. Third, and most important, there are empirically confirmed similarities between imagery in different modalities, such as isochrony (for more such similarities and details, see e.g., Jeannerod & Frak 1999, and references therein). If motor and visual imagery are subserved by different kinds of mechanisms, then such parallels have no obvious explanation. To the extent that it can be shown that similar information-processing structures underlie imagery regardless of modality, then commonalities such as isochrony have a natural explanation.

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 efference 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 *some* cases the motor aspect is minimal or absent, because the emulation required to support the imagery does not require efference copies (as when one imagines a vase of flowers), though of course in other cases it does (as in Wexler's studies). The mechanisms in both cases are the same. The difference is that the kinds of tasks emphasized in visual imagery often lack motor components, whereas motor imagery tasks nearly always highlight exactly the efference-copy driven aspects of the mechanism.

5. Perception

5.1. Sensation versus 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 upon, or suggested by, this input. So, for example, when looking at a wire-frame cube, the *sensory* input consists of twelve co-planar line segments: four horizontal, four

https://doi.org/10.1017/S0140525X04000093 Published online by Cambridge University Press

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 and surfaces, 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 an *environment emulator.* Clearly one of the primary functions of perception is the formation of an accurate estimate of the environment, and this will be embodied in the environment emulator.

Look again at Figure 7. In section 4, 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. In the feedback control scheme, the feedback is necessarily whatever signal is produced by the plant's sensors, and this imposes 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.

In 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 neither necessary nor, in many cases, desirable. The real process/plant has many state variables, only a small 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 amodal 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 a modality-specific measurement. There is no need to do this. The emulator is a neural system: any and all of its relevant states can be directly tapped.[13] 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 significant, because, as already mentioned, a modality-specific 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 out control signals 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 in 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. This is true both on the efferent and afferent ends. In the emulation theory, such systems are engaged in *sensation.* The more sophisticated systems have their goals set in terms of objects and locations in the environment, and receive information in terms of objects and locations in their environment. In the emulation theory, these systems are engaged in *perception.*

5.2. The environment 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 kinesthesis of the MSS, as done above in section 3. 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 and surfaces of various sizes, shapes, and egocentrically specified locations, entering into force-dynamic interactions with each other and the organism. Of course, the proposal that an organism's brain has the wherewithal to maintain models with these characteristics does not imply that such models are always complete or accurate. Indeed, there is reason to believe that such models are often schematic, incomplete, and inaccurate. But for my present purposes, the degree of completeness is not important. So long as it is agreed that brains are not always without any kind of model of their environment (and surely this much is unobjectionable), the question concerning what such models are and how they are implemented is unavoidable. How detailed, complete, and accurate such models are in what kinds of circumstances is a crucial issue, but not the current issue.

This involves a combination of *where, what* and *which* systems. The *what* and *where* systems are posited to be located in the ventral and dorsal processing streams respectively (Ungerleider & 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

https://doi.org/10.1017/S0140525X04000093 Published online by Cambridge University Press

the location of objects. In addition to these, a *which* system, whose task is indexing and tracking object identity through changes in locations and properties, also appears to be in play (Pylyshyn 2001; Yantis 1992).[14]

These systems comprise the core of the environment emulator in 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 of 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.

5.3. On perception and imagery

In contrast to traditional theories of perception that treat it as largely a bottom-up process driven entirely be the sensory periphery, Stephen Kosslyn has developed one of the most influential accounts of the nature of visual perception and its relation to visual imagery. Kosslyn's theory is complex and has many features I will not mention. But I want to focus for now on the main factor that, according to Kosslyn, accounts for the "top-down" nature of visual perception: visual imagery. Kosslyn maintains that imagery processes are used to aid perceptual processing, by filling in missing information on the basis of expectations, for example. 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." In fact, not only does Kosslyn claim that visual imagery is an ingredient in visual perception, but he also includes what he calls "motion added" imagery (Kosslyn 1994), which is imagery whose character is in part determined by off-line motor commands.

Although Kosslyn does not cast any of his points in terms of the emulation framework or Kalman filters, it should be clear that the emulation theory is particularly well-suited to explaining the information-processing infrastructure of Kosslyn's account. In the first place, the emulation framework explains exactly how it is that the same process that can operate off-line as imagery, can operate on-line as the provider of expectations that fill in and interpret bare sensory information. In this case, it is the a priori estimate that is then corrected by the sensory signal. Furthermore, the emulation framework's explanation of imagery as the driving of emulators by efferent motor areas in fact predicts what Kosslyn's theory posits, that in many cases the imagery that is used to aid in perception is the product, in part, of the activity of motor areas.

Striking confirmation of this view of the relation between motor-initiated anticipations and perception comes from a phenomenon first hypothesized by von Helmholtz (1910), and discussed and verified experimentally by Ernst Mach (1896). Subjects whose eyes are prevented from moving and who are presented with a stimulus that would normally trigger a saccade (such as a flash of light in the periphery of the visual field) report seeing the entire visual scene momentarily shift in the direction opposite of the stimulus. Such cases are very plausibly described as those in which the perceptual system is producing a prediction – an a priori estimate – of what the next visual scene will be on the basis of the current visual scene and the current motor command. Normally such a prediction would provide to specific areas of the visual system a head start for processing incoming information by priming them for the likely locations of edges, surfaces, et cetera. Just less than one hundred years after Mach published his experimental result, Duhamel, Colby, and Goldberg (Duhamel et al. 1992) published findings to the effect that there are neurons in the parietal cortex of the monkey that remap their retinal receptive fields in such a way as to anticipate immanent stimulation as a function of saccade efference copies. That is, given the current retinal image and the current saccade motor command, these neurons in the parietal cortex anticipate what the incoming retinal input will be. In particular, a cell that *will* be stimulated because the result of the saccade will bring a stimulus into its receptive field, begins firing in anticipation of the input, presumably on the basis of an efference copy of the motor command. (This is a neural implementation of the construction of an a priori estimate.)

While Kosslyn's theory of visual perception is very detailed, I hope that it is at least clear how these main aspects of his theory can be synthesized by the emulation framework into a yet broader account of the brain's information-processing structure, and, in particular, how the emulation framework makes clear the relation between Kosslyn's theory and specific results in cognitive neuroscience such as those of Wexler et al. (1998) and Duhamel et al. (1992).

5.4. 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 layout of the organism's environment from bare sensation. In my account, a good deal of this is represented in the KF-control diagram by the line that goes from the sensory residual to the 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 environment or the environment emulator. In the case of the amodal environment 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 emulation theory. The point is to show how *this* narrow process is part of a larger process, and to do 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 is one aspect of a complicated process that intimately involves motor control and imagery. I will address these in reverse order.

The minimal 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 (though,

https://doi.org/10.1017/S0140525X04000093 Published online by Cambridge University Press

as Kosslyn shows, some prominent theories do draw 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 (for a Kalman filter model of visual processing compatible with the emulation theory, see Rao & Ballard [1999]). Surely, to treat perception, imagery, and motor control as functionally distinct modules, as though any of them could do its job without the others, is to significantly distort the genuine neurophysiological 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. The conceptual significance of this is that it allows us to acknowledge the action/behavioral bias of perception without becoming anti-representationalists about perception.

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 completed percept is available. Rather, completed percepts of the environment are the starting point, in that the emulator always has a potentially self-contained environment emulator estimate up and running. This self-contained estimate is operational not only during imagery, but presumably also during dreaming (see Llinas & 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.*[15]

6. General discussion and conclusion

In this final section, I will make some *very rough* suggestions as to how the emulation theory of imagery and perception not only sheds new light on a number of issues in these fields, but also how it might synthesize aspects of other domains, such as reasoning and language, as well.

6.1. 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 typically 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 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 efference 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 ascription of a property to some object.

Nevertheless, such imagery is emphatically *unlike* a picture. This is difficult to appreciate largely because we typically, automatically and unconsciously, interpret pictures as having spatial/object import. But, strictly speaking, this import is not part of the picture. A picture or image, whether on a television screen, a piece of paper, or topographically organized early visual cortex, consists of a dimensionally organized placement of qualities, such as a pattern of colored pixels on a CRT. Seeing such a pattern as representing moving objects, such as Olympic sprinters in a race, involves interpreting the image. The picture itself has no runners, only pixels. 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.

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, in my account, sensation given an interpretation in terms of the amodal environment emulator. Therefore, although amodal imagery is not picture-like, it is not obviously sentential or propositional either. These amodal environment emulators are closely tied to the organism's sensorimotor engagement with its environment. The model is driven by efference copies, and transformations from one representational state to another follow the laws of the dynamics of movement and engagement (see Schwartz 1999), not of logic and entailment (as typically understood), or at

https://doi.org/10.1017/S0140525X04000093 Published online by Cambridge University Press

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 do not 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 as the emulation theory suggests, then this might go some way to explaining why two camps, one insisting on 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 as understood by the emulation theory, 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.

6.2. 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 (1983) 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 environment emulators are of exactly this sort.

Johnson-Laird's mental models, while arguably based on something like the representations made available through amodal 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 a 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 amodal space/object emulators, as understood in the current framework. If this is so, then the emulation framework might be a part of our eventual understanding of the relation between cognition and sensorimotor, perceptual, and imagery processes.

In a similar vein, Lawrence Barsalou (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 not only subserve 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 with the emulation framework, 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.

6.3. Other applications

There are a great number of other potential applications of the emulation framework in the cognitive and behavioral sciences. In a spirit of bold speculation, I will close by outlining a few of these.

6.3.1. Damasio's theory of reason and emotion. 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," implemented 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.

A payoff of understanding Damasio's proposal as a special case of the emulation framework in action is that it allows us to take Damasio's theory further than he takes it himself. 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 scheme for *emotional* perceptual processing. That is, just as perception of objects in the environment is hypothesized to involve a content-rich emulator-provided expectation that is corrected by sensation, so, too, might emotional perception involve a rich framework of 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 to yield the richer range of emotional perception that we gain as we age.

6.3.2. Theory of mind phenomena. 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

https://doi.org/10.1017/S0140525X04000093 Published online by Cambridge University Press

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, because 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 surrogate environment situation), and perhaps in their emotional situation with something like the emotion emulator discussed in the previous paragraph.

6.3.3. Situated robotics. Lynn Stein (1994) developed a robot, MetaToto, that uses a spatial emulator 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 to allow it to navigate more efficiently. By building up this map while exploring, MetaToto can then use this map 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.

6.3.4. Language. 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, are analogous to spatial/object representations posited here. Ronald Langacker's Cognitive Grammar framework (Langacker 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 Hoek (1995; 1997) has developed a very detailed account of pronominal anaphora within Langacker's Cognitive Grammar framework. Leonard Talmy (Talmy 2000), Lawrence Barsalou (Barsalou 1999; Barsalou et al. 1999), George Lakoff, and Mark Johnson (Lakoff 1987; Johnson 1987; Lakoff & Johnson 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 behavior and perception – exactly the sorts of representational structures made available by the various emulators described here.

6.4. 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 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 and articulate* 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, rather than to provide compelling data for its adoption. This synthesis is useful for a number for reasons. In the case of motor control, imagery, and perception, many researchers have assumed connections between these phenomena, but have not yet had the benefit of a single framework that details exactly how they are connected. And in the domains I have touched on even more superficially in sections 6.1–6.3, many researchers also make frequent appeal to their models being based in, or being continuous with, motor control, imagery, and perception, but again have not had the benefit of a synthesizing model that helps to make the nature of such connections perspicuous. While space has prevented going into these applications in detail, I hope that my brief remarks in these sections have at least made it clear in outline form how these "higher" brain functions might be synthesized with these other capacities.

In addition to this synthesizing potential, the emulation theory also manages to extend certain ideas in clearly useful directions, such as providing concrete ways of thinking of modal and amodal imagery and their potential interactions, and allowing more clarity regarding the mechanisms of motor imagery. (Even if one thinks that the simulation theory is correct, it is useful to have apparatus to clearly state it in such a way as to distinguish it from the emulation theory. Perhaps this can even lead to better experiment design.)

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 that has been robustly tested and largely vindicated. 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 when interpreting data or designing experiments. Perhaps when more researchers in a wider range of fields are familiar with the emulation theory and its potential applications, it will receive the kind of experimental attention that would be needed to determine the extent to which it is in fact used by the brain as I have claimed.

ACKNOWLEDGMENTS

I am grateful to the McDonnell Project in Philosophy and the Neurosciences, and the Project's Director Kathleen Akins, for financial support during which this research was conducted. I am also grateful to a number of people for helpful feedback on earlier versions of this paper, including Jon Opie, Adina Roskies-Davidson, Anthony Atkins, Brian Keeley, and Jonathan Cohen. A number of referees for this journal also provided extremely valuable suggestions and feedback that resulted in many improvements.

https://doi.org/10.1017/S0140525X04000093 Published online by Cambridge University Press

NOTES

1. Those interested in more technical details should see any of the many works that discuss KFs in detail, for example, Kalman (1960); Kalman and Bucy (1961); Gelb (1974); Bryson and Ho (1969); and Haykin (2001); for some discussion of applications of KFs (among other constructs) to understanding brain function, see Eliasmith and Anderson (2003).

2. 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.

3. See discussion in Hutchins (1995; especially Ch. 3).

4. What gets suppressed is the overt performance. Interestingly, however, a great many other bodily events normally associated with overt performance, such as increases in metabolic activity, heart rate, and so on, accompany many kinds of motor imagery. For a review, see Jeannerod (1994). In this article, when I speak of motor commands being suppressed in favor of the processing of an efference copy, I mean only the overt bodily movements are suppressed. It may even be so that in some cases there is a small degree of muscular excitation, perhaps because the motor signals are not completely blocked.

5. It should be noted that Johnson's position here is not exactly the same as Jeannerod's, because he claims that this imagery is used in order to construct a final motor plan. But, as far as I can tell anyway, Johnson nevertheless is maintaining that it is imagery that is being used, and that this imagery is the result of the "simulated" operation of efferent motor areas, those involved in planning a movement. The details are complex, though, and Johnson's position may not be a good example of what I call the simulation theory.

6. The situation here is complex. It is not clear to what extent and under what conditions the MSS emulator adapts as a function of plant drift. While the case of phantom limb patients suggests that it can, other cases of paralysis suggest that this is not always so (Johnson 2000b). I will simply note that the emulation theory itself need not take a stand on whether, and under what conditions, emulators are malleable. I use the example of apparent malleability in the case of phantom limb patients to make the contrast between the emulation and the simulation theories clear. But that clarificatory role does not depend on the empirical issue of the conditions under which such malleability actually obtains.

7. A benefit of the emulation theory over the simulation theory is that it allows us to make sense of the difference between (a) things which we cannot move but *do not feel paralyzed* and (b) things which we cannot move and do *for that reason* feel paralyzed. The first group includes not only our own body parts over which we have no voluntary control, such as our hair, but also foreign things such as other people's arms, chairs and tables, et cetera. We cannot move these things, but the phenomenology of their not being voluntarily movable is not like that of a paralyzed part. If mere lack of ability to produce a motor plan accounted for the feeling of paralysis, then all of these things should seem paralyzed. On the emulation theory, the feeling of paralysis is the product of a mismatch between a motor plan and the resultant feedback, whether from the body or the emulator. Such a mismatch is possible only when we can produce a motor plan that mismatches the result of the attempt to effect that motor plan.

8. One difference is that Murphy was an actual robot, whereas the model discussed here is completely virtual.

9. Mel's Murphy is a very simple system working in a very constrained environment. More complex environments, including those with objects that moved without the agent willing it, would be far less predictable. This is of course much of the reason why, in perception, the Kalman gain is set fairly high. I use Murphy because its simplicity makes it a good exemplar for introducing the basic ideas, and I simply note that real perceptual situations will require much more sophistication. For an example of the sort of complexities that a full version of this sort of mechanism would need to deal with, see Nolfi and Tani (1999) and the references therein.

10. The emulation theory predicts exactly Wexler et al.'s results. Thus, if it were the case that motor areas were not active during "active" visual imagery, or if it were the case that the specific nature of the motor command associated with the imagined movement (rotate right vs. rotate left, for example) were not recruited during such imagery, then this would be prima facie evidence against the emulation theory.

11. I call this kind of imagery *amodal* rather than *multimodal* because this sort of imagery, if it in fact exists, is not tied to any modality. But in a sense it is multimodal, because it can be used to produce a modal image in any modality so long as a measurement procedure appropriate to that modality is available. The expressions *amodal* and *multimodal* are used in many ways, and it may be that what I here am calling amodal imagery might be close to what some researchers have called multimodal imagery.

12. For example: If x is between a and b, and b is between a and c, is x necessarily between a and c? 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 needed, though of course it might be involved in specific cases.

13. What I have in mind here is the idea that the neurally implemented emulator represents states by things like firing frequencies, phases, and such. A neural pool that is representing the presence of a predator behind the rock by firing rapidly can be "directly measured" in the sense that other neural systems can be wired such as to sniff that pool's activation state, and hence be sensitive to the presence of the predator. A "measurement" of this state would yield a visual image of a rock, because the predator is not in the visual image, and hence the narrowly modal emulator would throw away relevant information.

14. Exactly how to understand such a system is not trivial. Understanding the *which* system as an attentional tagging mechanism is sufficient for present expositional purposes, but my suspicion, which I am not prepared to argue for here, is that it is a system that has richer representational properties, such as the constitution of basic object identity. Of course, the richer sort of mechanism, if there is one, will surely be based at least in part upon a simpler attentional tagging mechanism.

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

Open Peer Commentary

Redundancy in the nervous system: Where internal models collapse

Ramesh Balasubramaniam

Sensory Motor Neuroscience, School of Psychology, University of Birmingham, Edgbaston, B15 2TT, United Kingdom.

r.balasubramaniam@bham.ac.uk **http://www.bham.ac.uk/symon**

Abstract: Grush has proposed a fairly comprehensive version of the idea of internal models within the framework of the emulation theory of representation. However, the formulation suffers from assumptions that render such models biologically infeasible. Here I present some problems from physiological principles of human movement production to illustrate why. Some alternative views to emulation are presented.

In the target article, Grush presents a unified theory for psychology based on the idea that the nervous system uses internal mod-

https://doi.org/10.1017/S0140525X04000093 Published online by Cambridge University Press