els and evidence of redundant control inherent in many biological systems justify this approach. The next step from this perspective is to elaborate each process in the model in disparate contexts, a task receiving somewhat preliminary treatment in the article because of space limitations. Two of the processes are discussed here with respect to movement coordination.

Perhaps unsurprisingly, simple feed-forward and feedback models of motor control have often (though certainly not exclusively) been tested using single-limb, point-to-point movements. The control of even these relative "simple" actions is far from trivial, but the function of each component in Grush's control scheme is relatively straightforward in this case. But what of more complex, multilimb actions that implicitly entail coordination per se as part of the goal? Examples range from maintaining a 180-degree interlimb phase offset when beating a drum or when walking, to generating the constantly changing (but determinate) phase of the notes constituting a jazz riff on piano. The model still potentially applies in this case, but the processing demands of some model components are naturally increased. This is particularly the case for the measurement matrix in the "plant" (i.e., the record of appropriate sensory state variables) and the process model in the emulator.

Consider, first, the appropriate sensory variables constituting the measurement matrix for the state of the plant during coordinated actions. Different classes of sense organs provide, to lesser and greater degrees of accuracy, unidimensional information regarding muscle length and change in length, joint angle, muscle tension, and so forth. A direct measure of phase offset is obviously not available. Discrete point estimates of relative phase for drumming at a constant rate could be determined by monitoring the consistency of the half-periods between bilateral contact points, for example. But a *continuous* measure of phase lag, which is required to exploit the rapid feedback afforded by the model, requires substantially more processing; determining the phase offset between the end points of multisegmented effectors requires still more processing; and so forth. The problems are not intractable, but transforming the raw information into a form relevant to the task at hand presumably requires time and is subject to error.

Now consider the complexity of the process model that is needed to emulate a coordinated action. A number of mappings reflecting the mechanics associated with single-limb control must be included, comprising such considerations as the length-tension and force-velocity properties of muscle, gravity, passive torques interacting nonlinearly among segments, and so forth. When additional limbs are involved in the action, the need to consider so-called "coordination dynamics" is invoked. As Kelso and colleagues noted some time ago (e.g., Haken et al. 1985), concurrent limb movements are influenced not only by a central control signal but often by nonlinear coupling between the effectors as well. This coupling can result in (initially) unpredictable behavior. A transition to a completely unintended phase offset between the limbs can occur if the drive is scaled past a critical level, for example. The timing demands of the effectors clearly influence interlimb interactions, but spatial interference is evident as well (e.g., Walter et al. 2001). The dynamics of these space-time interactions are part of more global plant dynamics, so the model is not invalidated. The obvious point is simply that the process model for human coordination is substantially more complex than, for example, the mapping between the signal driving torque motors and the resultant motion of a robot arm. But the more insidious question that emerges when attempting to fill in the details of this model component is whether the complexity of the emulation process is amenable to our attempts to capture it with an elegant, articulated account.

Although presented separately, the twin concerns elaborated above are of course related. The Kalman gain is only as effective as the rapidity and accuracy with which the actual sensory signal and the mock sensory signal are generated. Illuminating these two processes alone poses an interesting challenge in further developing the model to account for complex, coordinated actions.

Representing is more than emulating

Hongbin Wang[a] and Yingrui Yang[b]

[a]*School of Health Information Sciences, University of Texas Health Science Center at Houston, Houston, TX 77030;* [b]*Department of Cognitive Science, Rensselaer Polytechnic Institute, Troy, NY 12180.*

Hongbin.Wang@uth.tmc.edu **yangyri@rpi.edu**

http://www.shis.uth.tmc.edu/

Abstract: Mental representations are more than emulations. Different types of representations, including external representations, various mental models (distorted and abstract), and emulative models, can all play important roles in human cognition. To explain cognitive performance in a specific task, a systematic analysis of the underlying representational structures and their interactions is needed.

The essence of the target article is that the human brain maintains various internalized and adaptive representations of the body and the environment, which, via Kalman filter–like control schemes, can emulate/simulate/mimic quite precisely the actual brain-body and human-environment interactions. Although the idea of mental representations as emulations is intuitively appealing and theoretically interesting, we would like to comment on the adequacy of the theory and argue that mental representation is more than emulation.

Few will doubt that a mental emulator would be useful if one exists. As demonstrated extensively in the target article, an emulator implements an essential look-ahead capacity and therefore permits a systematic treatment for a wide variety of cognitive phenomena, including motor and visual imagery, perception, reasoning, and language. However, the real issue is whether these powerful mental emulators do exist in the brain and in the mind. We would expect that an emulation theory of representation would provide answers to important questions such as why mental emulators are necessary and how they are acquired. Unfortunately, although the target article provided a detailed account of how mental emulators help to implement various cognitive functions, it has left these fundamental questions largely unanswered. The author has felt reluctant to give evidence that *counted against* the theory and admitted that the goal was not "to provide compelling data for its adoption" (sect. 6.4, para. 1). The result is a less convincing theory.

In this commentary we raise two issues that we think are directly related to the theory's treatment of mental representations as emulations. The first one has to do with the relationship between emulation and other types of representations. Numerous studies in the broad field of cognitive psychology and cognitive science have shown that people can and do use different types of representations in performing various cognitive tasks (Donald 1994; Palmer 1978; Rumelhart & Norman 1988). To explain cognitive performance in a specific task, an analysis of the underlying representational structures and their interactions is needed. For example, Zhang and colleagues have systematically studied the nature and function of external representations (Zhang 1997; Zhang & Norman 1994). According to this view, certain types of information, such as affordance and salient spatial relationships, exist as external representations in the world and can be directly perceived and used by the mind. As a result, for these types of information no mental internalization is needed to support actions. In one of the studies, Zhang et al. found that the subjects' performance in solving different isomorphs of tic-tac-toe was determined by the interplay of different types of representation and that, when more information was available directly through external representations, the task was easier.

Although various internal representations are an important type of representations, they are typically compressed, segmented, and distorted forms of the represented entity but not emulations. These characteristics result from the mind's nontrivial reorganization of sensory input and are constrained by the limited processing power of the brain. One example of this is how the mind represents space (Tversky 2000). Recent evidence in cognitive neuroscience has shown that the "where" pathway is more complex than a unified object-location emulator. It consists of multi-

https://doi.org/10.1017/S0140525X04470095 Published online by Cambridge University Press

ple spatial centers, each processing and extracting distinct spatial information. As a result, space is represented in the mind not once but multiple times, not in a unified manner but segmented. Each representation is a (distorted) salience map with distinctive frames of reference (see Wang et al. 2001).

While we agree that emulation is yet another type of mental representation, we maintain that it is the goal but not the means. After all, when we have an emulator of the universe, to a certain degree we have understood how it works (see also Lewontin 2001 who suggests that the essence of science is to seek metaphor/emulation). In early development children are capable of interacting with the environment, but it is implausible to assume that they possess an emulator of the environment. Learning to interact with a model/emulation of the environment is certainly desired, but children must start with a model-less interaction mode and gradually learn to acquire such a model (e.g., Sutton & Barto 1998).

We would like to clarify our argument further by raising the second issue, which is related to the author's attempt to use amodal environment emulation to explain Johnson-Laird's mental models in human reasoning. First of all, according to Johnson-Laird (e.g., Johnson-Laird 2001), each mental model represents a possibility, and people by default represent only what is true but not what is false. In this sense each mental model is a fragmented representation of what is true, and constructing mental models involves considerable information processing at the semantic level. Treating mental models as emulations raises questions about how the emulators are smart enough to perform these semantic processes. Therefore, it seems that the claim does not solve the problem but simply gives it a different name.

In addition, treating mental models as space/object emulations against sentential representations would encounter difficulty in representing certain types of ordinary inferences that involve compound statements (Yang & Bringsjord 2003). The current mental model theory is powerful in representing prepositional reasoning but it has limited power in quantified predicate reasoning. Mental models theory provides initial models for truth connectives such as disjunction, and provides initial mental models for quantified atomic sentences such as: "(For all x) Ax." In combination, it can represent a compound statement such as: "(For all x) Ax or (for all x) Bx." It is difficult, however, for the current mental models theory to have a spatial/object style presentation for a quantified compound statement such as: "(For all x)(Ax or Bx)." For statements of this type, Yang and Bringsjord (2003) argue that some special syntactic structures would have to be constructed. Though one may still find ways to represent these mental models as space/object emulations (e.g., $[x]^{Ax}_{Bx}$, as a quantified disjunction), they apparently make mental models less efficient. At this point, the emulation hypothesis would face a philosophical question: Does cognition need to be maximally efficient?

Small brains and minimalist emulation: When is an internal model no longer a model?

Barbara Webb

Institute for Perception, Action and Behaviour, School of Informatics, University of Edinburgh, Edinburgh EH1 2QL, United Kingdom.
bwebb@inf.ed.ac.uk
http://www.informatics.ed.ac.uk/people/staff/Barbara_Webb.html

Abstract: Many of Grush's arguments should apply equally to animals with small brains, for which the capacity to internally model the body and environment must be limited. The dilemma may be solved by making only very approximate predictions, or only attempting to derive a "high-level" prediction from "high-level" output. At the extreme, in either case, the "emulation" step becomes trivial.

Grush makes a good case that concepts of emulation from control theory can illuminate a range of issues when applied to the brain. But this approach should apply to a wider range of brains than the human ones that form the focus of the article. Animals with small brains, such as insects, face many of the same behavioural challenges that the emulation theory is proposed to solve. For example, flies need to distinguish the visual slip that results from self-initiated steering, from that caused by external disturbances if they are to use the latter for stabilisation. It was just this sort of problem that first led to the explicit formulation of the concept of efference copy by von Holst and Mittelstaedt (1950; simultaneously described by Sperry 1950 as "corollary discharge").

Grush notes, in the field of motor imagery, a surprising lack of appreciation of the fact that an internal copy of the motor output is not in itself sufficient to explain the prediction of sensory experience. This failure to deal with the problem of the difference between the format of the output and input signals is pervasive in the biological literature on efference copy. Von Holst and Mittelstaedt, in their early formulations, describe efference copy as being compared to the sensory reafference like a "photo-negative," without any discussion of the potential differences between an "image" of the motor output and the "image" resulting from sensory input. In principle, for the example of the fly, it would be necessary for the fly to model the exact dynamics of how muscle commands affect wing movements, how wing movements turn the body, how the turning body affects the visual input, and how the visual sensors will respond to this input, to be able to produce a prediction that is spatio-temporally accurate. Indeed, as the response of its visual system to image velocity is known to be highly dependent on the contrast and spatial frequency of the scene, the fly must in theory have advance knowledge of the scene it will encounter as it turns if it is to calculate an exact prediction. It seems unlikely that all this is actually occurring in the small brain of the fly.

There are several ways out of this problem. The first is the recognition that the emulation may not be complete and accurate (briefly referred to by Grush in sect. 5.2), and that it does not need to be so to be useful. A simple scaling of the strength of the turn signal might be a good enough approximation of the size of the resulting optomotor signal for the fly. An issue that then arises is the following: How "schematic" (to use Grush's term) can the emulator be yet still count as an emulator? If it is defined by its role in the control cycle, as pictured in Figure 4 of the target article, then arguably even a simple inhibitory signal sent from motor areas to sensory areas should count, such as the "saccadic suppression" found in locusts (Zaretsky & Rowell 1979). On the other hand, it does not seem warranted to say that the animal is using an "internal model" in such a case. But what additional detail in the emulation would justify the ascription of a model? What if the prediction is also direction-specific, as for the *Drosophila* optomotor response (Heisenberg &Wolf 1984)? Or can be retuned online to be proportional in size (Mohl 1988)? Or is well matched in its time course, as in the inhibitory signal anticipating self-produced sound bursts in the cricket (Poulet & Hedwig 2003)? Must it estimate more than one dimension of the input signal? Or is the defining characteristic of a true internal model that it can also be run offline for planning? One might think this would eliminate examples from "small-brained" animals, although there is evidence for route-planning in jumping spiders (Tarsitano & Andrew 1999).

A second way out of the problem of having to emulate all the details of the motor system, environment, and sensors is discussed in more detail by Grush in section 5.1. The idea is to have the emulator predict higher-level representations (such as the new layout of objects in the world relative to the observer after a particular movement), rather than the raw sensations this new layout would produce. In the case of the fly, the prediction might be supposed to correspond to the output of the higher-level neurons that integrate motion across the visual field, rather than the response of each elementary motion detector. Of course, this shortcut to avoid modelling peripheral mechanisms can apply on the motor output side as well: The emulator can take as input a high-level motor command such as "rotate ten degrees left" and assume that this

https://doi.org/10.1017/S0140525X04470095 Published online by Cambridge University Press