

"Embodiment and the Philosophy of Mind" in A. O'Hear (ed) CURRENT ISSUES IN PHILOSOPHY OF MIND: ROYAL INSTITUTE OF PHILOSOPHY SUPPLEMENT 43 (Cambridge University Press:1998) P. 35-52. To be reprinted in Alberto Peruzzi (ed) MIND AND CAUSALITY (Holland, forthcoming 2003). Appears in Spanish as "Incorporizacion y la filosofia de la mente" in P. F.. Martinez-Freire (ed) FILOSOFIA ACTUAL DE LA MENTE. CONTRASTES: SUPPLEMENTO 6 (2001)

EMBODIMENT AND THE PHILOSOPHY OF MIND

Andy Clark

*Philosophy/Neuroscience/Psychology Program
Department of Philosophy
Washington University
St. Louis, MO 63130
e-mail: andy@twinearth.wustl.edu*

1. INTRODUCTION: THE REDISCOVERY OF THE BODY AND OF THE WORLD.

Cognitive Science is in some sense the science of the mind. But an increasingly influential theme, in recent years, has been the role of the physical body, and of the local environment, in promoting adaptive success. No right-minded Cognitive Scientist, to be sure, ever claimed that body and world were completely irrelevant to the understanding of mind. But there was, nonetheless, an unmistakable tendency to marginalize such factors: to dwell on inner complexity whilst simplifying or ignoring the complex inner-outer interplays that characterize the bulk of basic biological problem-solving¹. This tendency was expressed in, for example, the development of planning

¹Some exceptions to this trend include the work such Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton-Mifflin, Merleau-Ponty, M. (1942). *La Structure du Comportement* (A. Fisher, Trans.). France: Presses Universitaires de France. Work in Animate Vision and ecological optics (see Section 2 below) is clearly influenced by Gibsonian ideas, while more philosophical treatments (such as Varela, F., Thompson, E., & Rosch, E.

algorithms that treated real-world action as merely a way of implementing solutions arrived at by pure cognition (more recent work, by contrast, allows such actions to play important computational and problem-solving roles²). It was also expressed in David Marr's³ depiction of the task of vision as the construction of a detailed three-dimensional image of the visual scene. For possession of such a rich inner model effectively allows the system to "throw away" the world and to focus current computational activity in the inner model alone⁴. More generally, the whole vision of cognition as inner operations on internal world models reflects an explanatory strategy which might reasonably be dubbed isolationism⁵:

(Isolationism)

The world is (just) a source of inputs and an arena for outputs. And the body is just an organ for receiving inputs and effecting outputs (actions). The task of early processing is to render the inputs as an inner world-model of sufficient

(1991). *The Embodied Mind*. Cambridge, MA: MIT Press) explicitly acknowledge Merleau-Ponty. There is a brief discussion of these historical roots in Clark (1997), Chapter 8.

²See e.g., Agre, P., & Rosenschein, S. (Eds.). (1996). *Computational Theories of Interaction & Agency*. Cambridge: MIT Press, Kirsh, D., & Maglio, P. (1995). On Distinguishing Epistemic from Pragmatic Action. *Cognitive Science*, 18, 513-549, Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA: MIT Press.

³See Marr, D. (1982). *Vision*. San Francisco, CA: W.H. Freeman.

⁴See Ballard, D. (1991). Animate Vision. *Artificial Intelligence*, 48, 57-86, Churchland, P., Ramachandran, V., & Sejnowski, T. (1994). A Critique of Pure Vision. In C. Koch & J. Davis (Eds.), *Large-Scale Neuronal Theories of the Brain*. Cambridge, MA: MIT Press.

⁵Roboticians refer to this isolationist vision as the (increasingly discredited) idea of a simple Sense-Think-Act Cycle. See e.g., Malcolm, C., Smithers, T., & Hallam, J. (1989). *An Emerging Paradigm in Robot Architecture*. Edinburgh University Department of Artificial Intelligence.

thickness to allow the bulk of problem-solving activity to be defined over the inner model alone.

Isolationism, it is fair to say, is in increasing disrepute. But the precise shape of an alternative paradigm remains unclear. Anti-isolationist assertions range from the relatively innocent insistence that we won't achieve a balanced vision of what the brain does until we pay more heed to the complex roles of body and world, to the self-consciously revolutionary accusation that mind itself is not, after all, a special realm populated by internal models and representations so much as an inextricable interwoven system, incorporating element of brain, body and world -- a system which resists informative analysis in terms of the old notions of model, representation and computation⁶. The most radical anti-isolationist vision thus depicts human beings as a species of (so-called) *post-Cartesian agent*⁷. The post-Cartesian agent is a locus of

⁶See Haugeland, J. (1995). Mind Embodied and Embedded. In Y.-H. Houn & J.-C. Ho (Eds.), *Mind and Cognition* (pp. 3-38). Taipei, Taiwan: Academia Sinica, van Gelder, T. (1995). What Might Cognition Be, If Not Computation? *Journal of Philosophy*, XCII(7), 345-381, Port, R., & Gelder, T. V. (Eds.). (1995). *Mind as Motion: Dynamics, Behavior, and Cognition*. Cambridge, MA: MIT Press, Thelen, E., & Smith, L. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA: MIT Press, Varela, F., Thompson, E., & Rosch, E. (1991). *The Embodied Mind*. Cambridge, MA: MIT Press.

⁷This vision is clearly contemplated in e.g., Haugeland, J. (1995). Mind Embodied and Embedded. In Y.-H. Houn & J.-C. Ho (Eds.), *Mind and Cognition* (pp. 3-38). Taipei, Taiwan: Academia Sinica and van Gelder, T. (1995). What Might Cognition Be, If Not Computation? *Journal of Philosophy*, XCII(7), 345-381, though both authors recognize the large space of intermediate possibilities. The term "Post-Cartesian Agent" is from van Gelder, T. (1995). What Might Cognition Be, If Not Computation? *Journal of Philosophy*, XCII(7), p. 381. See also Haugeland, J. (1995). Mind Embodied and Embedded. In Y.-H. Houn & J.-C. Ho (Eds.), *Mind and Cognition* (pp. 3-38). Taipei, Taiwan: Academia Sinica, p. 36, Thelen, E., & Smith, L. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA: MIT Press, p. 338, Port, R., & Gelder, T. V. (Eds.). (1995). *Mind as Motion: Dynamics, Behavior, and Cognition*. Cambridge, MA: MIT Press, p. IX.

knowledge, acts for reasons and has beliefs and desires. Yet she harbors no internal representations and resists analysis in terms of any cognitively important distinctions between “inner” and “outer” processes, between perception, cognition & action, or between mind, body, and world.

I shall argue that the post-Cartesian vision is unconvincing and that a key move in the argument (a move I dub the “Cognitive-to-Coping Shift”) is both dialectically suspect and empirically unsound. More positively, I shall suggest that a weaker anti-isolationist stance still requires some deep revisions in our understanding of the inner vehicles, the contents and the adaptive roles of internal representation and inner world models. The foundational and conceptual challenges are real enough, even when stripped of their radical post-Cartesian trimmings.

2. INNER SYMBOL FLIGHT.

The outright rejection of the notion of internal representation is usefully seen as the extreme limiting case of a (generally admirable) process of inner symbol flight. This process involves the progressive rejection of more and more of the apparatus and assumptions associated with the vision of cognition as the manipulation of chunky inner symbols. According to this simple (and historically important) vision, semantically sensible transitions between mental states are explained in terms of syntactically constrained transitions between inner symbol strings. These symbol strings contained discrete elements corresponding rather closely to the semantic elements identified in sentential descriptions of the relevant mental states. Thus, the thought that John loves Mary is realized as a complex inner symbol string that incorporates distinct and

independently manipulable elements standing for 'John' 'loves' and 'Mary'⁸.

⁸See Fodor, J., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28, p. 13.

This vision of simple inner symbolic atoms (unstructured base items) corresponding rather closely to familiar concepts and relations enshrined in daily discourse was challenged by the development of distributed connectionist⁹ models. The ‘sentential paradigm’¹⁰ was replaced, in this research, by a vision of internal representations as distributed patterns of activity across a whole array of simple processing units. Such distributed patterns were allowed to overlap in semantically significant ways, giving rise to a variety of computationally significant side-effects including free generalization, damage-resistance, etc.¹¹.

More recently still, we have witnessed increased attention to the temporal dynamics of the inner representational vehicles. The use of e.g., single recurrent neural networks¹² allows information to be encoded not just in instantaneous patterns of activity but in temporally extended processing trajectories. In these networks, much of the information-processing power resides in the way a current state allows or restricts future change and evolution. The progression has thus been from a view of simple, atomistic inner symbols, to a notion of spatially distributed patterns, to a notion of spatially and temporally distributed patterns. The inner vehicles of content, courtesy

⁹See Rumelhart, D., & McClelland, J. (1986). On learning the past tenses of English verbs. In D. Rumelhart & e. al. (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Vol. 2, pp. 216-271). Cambridge: MIT Press..

¹⁰P.S. Churchland (1986).

¹¹See Clark, A. (1989). *Microcognition: Philosophy, Cognitive Science and Parallel Distributed Processing*. Cambridge: MIT Press for discussion.

¹²Elman, J. (1991). Representation and structure in connectionist models. In G. Altman (Ed.), *Cognitive Models of Speech Processing*. Cambridge, MA: MIT Press.

of this progression, have come to look less like simple inner *states* and more like complex inner *processes*.

The metamorphosis, moreover, is probably still incomplete. Some rather plausible next steps include seeing the inner vehicles as multiply functional and seeing the inner architecture as dynamically reconfigurable. Multiple functionality would mean that one and the same inner resource may play a variety of content-bearing roles¹³ (perhaps varying in accordance with local content). Dynamic reconfigurability would mean that the inner structures are themselves subject to rapid change and reorganization, as when the release of a chemical neuromodulator causes two neural networks to temporarily ‘fuse’ and behave as one.

The moral, then, is that our understanding of the nature of the (putative) inner vehicles of content is in a state of extreme flux, characterized by a rapid flight from the initial image of static, chunky unstructural inner symbols. This flight has, in addition a more content-related aspect. For at the same time as the inner vehicles become more complex, so the characteristic contents have become more partial and fragmentary. This is because the emphasis has shifted from isolationist forms of problem-solving towards iterated series of agent-environment interactions. This shift lies at the very heart of the agenda of a more embodied and environmentally embedded approach to cognitive science and is nicely exemplified by recent work in the field known as

¹³For some hints of such content-sensitive complexity, see Knierim, J., & VanEssen, D. (1992). Visual Cortex: Cartography, Connectivity and Concurrent Processing. *Current Opinion in Neurobiology*, 2, 150-155.

Animate Vision¹⁴.

Recall Marr's depiction¹⁵ of the task of vision. The task, according to Marr is to construct a rich inner model of the three dimensional visual scene on the basis of the available (two dimensional) input information. Work in Animate Vision, by contrast, depicts the task as, simply, the use of visual strategies to control behavior, in real-world contexts, at as low a computational cost as possible. To this end, Animate vision avails itself of three central ploys.

- 1). The use of task-specific cues and shortcuts.
- 2). The use of body-centered (egocentric) strategies.
- 3). The use of repeated environmental interactions.

Task-specific cues and shortcuts include, for example, the use of 'personalized' idiosyncratic strategies such as searching for bright yellow (a cheap, easy visual cue) when searching for *my* coffee cup (which is canary yellow). Egocentric strategies include the use of so-called "deictic pointers" (see below). And repeated environmental interactions allow us, for example, to visit and re-visit different aspects

¹⁴Ballard, D. (1991). Animate Vision. *Artificial Intelligence*, 48, 57-86.

¹⁵Marr, D. (1982). *Vision*. San Francisco, CA: W.H. Freeman.

of a visual scene retrieving specific information only as and when required.

The case of deictic pointers merits a longer look. A pointer in classical Artificial Intelligence, is an inner state that can function in self-contained computational routines but which can also “point to” other data structures¹⁶. This pointing allows the retrieval, when required, of more detailed information, and the effective binding of certain items of information to others. Such binding can be temporary, as when we bind certain features (e.g., bright yellow) to certain current visual locations (top left of visual field).

Deictic pointers, however, are not inner markers but actual bodily orientations (such as saccadic eye movements) that play the same kind of particular aspect of a visual scene, we may both retrieve more detailed information and achieve a kind of temporary variable binding in which we associate the detected features with a given spatial location. A further example is the binding of a reaching-and-grasping routine to a target object using what is informally called a “do-it-where-I’m-looking” strategy. Here, the system is set up so that the grasping motion is directed to the fixated visual location. In all these cases:

The external world is analogous to computer memory. When fixating a location, the neurons that are linked to the fovea refer to information computed from that

¹⁶See e.g., Pylyshyn, Z. (Ed.). (1987). *The Robots Dilemma: The Frame Problem in Artificial Intelligence*. Norwood: Ablex.

location. Changing gaze is analogous to changing the memory reference in a silicon computer.

D. Ballard, M. Hayhoe, P. Pook & R. Rao
“Deictic Codes for the Embodiment of
Cognition” *Behavioral & Brain Sciences* (in
press), p. 6.

The general thrust of the Animate vision research, then, is that bodily actions (such as saccadic eye motions) play vital computational roles, and that repeated agent-environment interactions obviate much of the need to create all-purpose, detailed internal world models. Instead, we visit and re-visit different aspects of the scene as and when required, allowing the world to function as “its own best model”. The research program is this staunchly anti-isolationist. But it is not by any means ‘post-Cartesian’ -- it does not reject the ideas of internal models and representations, so much as reconfigure them in sparse and more interactive image. We thus read of ‘inner databases’ that associate objects (e.g., my car keys) and locations (on the kitchen table), of ‘internal featured representations’, ‘indexical representations’, and so on. What is being rejected is this not the notion of inner content-bearing states per se, but rather the much stronger notion of rich, memory-intensive, all-purpose forms of internal representation.

A similar conclusion is suggested by work¹⁷ in real-world robotic navigation, in which, knowledge of location is encoded as a perceptio-motor routine that will actually

¹⁷Mataric (1991). This work is further discussed in Clark (1997), Chapter 2.

move the robot from its present position to the desired spot. In this way, the inner map is itself the controller of the appropriate action. There is no need for a further system to access the 'map' and to plan a route. Instead, the knowledge is at once both descriptive and prescriptive¹⁸ -- a dual nature that affords great economies both in terms of response-time and computational effort.

The crucial distinction, it seems to me, is thus not between representational and non-representational solutions so much as between action-neutral forms of internal representation (which may increase flexibility but require additional computational work to specify a behavioral response) and action-oriented forms (which build the response into the representation itself).

The best work in Animate vision and real-world robotics, I claim, suggests *at most* that the use of truly action-neutral internal representation may be rather rare in biological cognition (my own view discussed briefly in Section 5 below, is that the use of such representations coincides, rather exactly, with the possession of a rich public language). Such conclusions are radical and challenging. But they fall well short of a full post-Cartesian rejection of inner models and representations. What considerations might drive us to question the idea of inner content-bearers *tout court*?

¹⁸For more on this theme, see Millikan, R. (to appear). Pushmepullyou Representations. In L. May, M. Friedman, & A. Clark (Eds.), *Mind & Morals*. Cambridge: MIT Press.

3. RADICAL INTERACTIONISM.

One leading anti-representationalist argument¹⁹ turns on the presence of dense, reciprocal causal exchanges uniting agent and environment in a complex web of mutual influence. Under such conditions, it is argued, the kind of de-composition and analysis that works so well in the case of e.g., a contemporary computer program simply gets no foothold. The problem (it is suggested) is that the notion of x representing y is too one-way and too simplistic to do justice to cases in which x is continuously affecting and being affected by y and vice versa. Yet typical agent-environment interactions often present just such a complex and circular causal profile.

Consider ballroom dancing. As you dance, your motions (if you are a good dancer!) are both continuously influenced by and influence upon, those of your partner: the two sets of motions “co-evolve” in a highly interdetermined way. Not is the presence of two human agents essential to the phenomenon. The same holds true as you windsurf: you constantly affect and are affected by the set of your rig. Van Gelder²⁰ makes the same point using the example of the Watt (or centrifugal) governor -- a device which maintains a steam engine at a steady speed by both affecting and

¹⁹This argument is the centerpiece of van Gelder, T. (1995). What Might Cognition Be, If Not Computation? *Journal of Philosophy*, XCII(7), 345-381 and is also visible in van Gelder, T., & Port, R. (1995). It's About Time: An Overview of the Dynamical Approach to Cognition. In R. Port & T. v. Gelder (Eds.), *Mind as Motion: Explorations in the Dynamics of Cognition* (pp. 1-44). Cambridge, MA: MIT Press, Thelen, E., & Smith, L. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA: MIT Press, Varela, F., Thompson, E., & Rosch, E. (1991). *The Embodied Mind*. Cambridge, MA: MIT Press. Other anti-representationalist arguments are considered in Clark (1997).

²⁰van Gelder, T. (1995). What Might Cognition Be, If Not Computation? *Journal of Philosophy*, XCII(7), 345-381. This example is treated in detail in Clark, A., & Toribio, J. (1994). Doing Without Representing? *Synthese*.

being affected by the engine speed. Such episodes of mutual influence were much discussed both in early cybernetics²¹ and in the work of the french phenomenologist Maurice Merleau-Ponty²².

²¹For example, W. Ross Ashby's *Introduction to Cybernetics* (Wiley, 1956).

²²Maurice Merleau-Ponty, *The Structure of Behavior* (Beacon, 1963). Originally *La Structure du Comportment* (Presses Universitaire de France, 1942)

Where such continuous, dense, circular causal influence obtains, it is argued, the tools of representational (and computational) analysis run aground. The idea of explaining the shape of the on-going agent-environment interaction by depicting an inner state as representing an outer one is coarse and distortive. Instead, inner and outer co-evolve in a mathematically precise way that is best captured (so the argument goes) by the use of coupled differential equations in which the current value of certain internal variables appear as parameter setting in the evolution equation for the external system and vice versa²³. Fortunately, the details of such a dynamical systems model are unimportant for present purposes²⁴. What matters is rather the

²³For an accessible introduction to these dynamical approaches, see Kelso (1995). A classic text is Abraham, R., & Shaw, C. (1992). *Dynamics -- The Geometry of Behavior*. Redwood, CA: Addison-Wesley.

²⁴For a fuller discussion, see Clark (1997), Chapters 5, 6, and 8.

general shape of the argument. Van Gelder²⁵ puts it well;

The internal operation of a system interacting with an external world can be so subtle and complex as to defy description in representational terms

Ibid, p. 381.

Before responding to this argument, it is worth pausing to clarify the challenge. For what is at issue is not the status of certain systems (ourselves, for example) as *representers*. That is a given. We surely do represent our world, our past, our possible futures, our absent friends and so on. We think of these things and states of affairs and thus represent them to ourselves. What is not a given (and what is at issue here) is that we use *internal representations* to do so. The point (and I think it is a good one) is that the notion that cognition involves internal representations (and computations defined over them) is meant to be not a simple rehearsal of the fact that we are thinkers, but a substantial and explanatorily potent empirical hypothesis: the kind of thing that could indeed turn out to be false. The claim, to a first approximation, is that there are distinct, identifiable inner resources whose systemic or functional role is to stand in for specific features or states of affairs.

²⁵van Gelder, T. (1995). What Might Cognition Be, If Not Computation? *Journal of Philosophy*, XCII(7), 345-381.

This notion of internal stand-ins is however, itself ambiguous. It is ambiguous²⁶ between a weak notion, in which x ‘stands in’ for y if x is an inner resource that a) carries information about y and b) is used to control behavior, and a strong notion in which the inner resource must be capable of functioning as a genuine surrogate, i.e., be capable of systematically controlling appropriate behavior even if y is absent or non-existent. A neural population²⁷ closely keyed to bodily orientation and used to control skilled action may thus be counted as a kind of weak stand-in. And even this weakly representational gloss tells us something useful about the purpose of the neuronal population and may shed light on larger scale systemic organization (we may see which other neuronal populations access that specific body of information and hence gain insights into their roles). But such a population, though it engages in information-based control of action, need not be capable of driving appropriate actions in the absence of the (weakly represented) state of affairs. It is this capacity to act as

²⁶See Clark, A., & Grush, R. (1997). *Towards a Cognitive Robotics* (PNP Technical Report No.). Washington University in St. Louis..

²⁷For example, the posterior parietal neuronal population in the rat which encodes information about which way the rat’s head is facing and which is exploited in radical maze running -- see McNaughton & Nadel (1991).

an inner surrogate in the absence of the target environment feature that I suggest, characterizes the strongest and most conceptually unequivocal cases of internal representation²⁸.

²⁸David Israel (Bogdan on Information, *Mind & Language*, vol 3:2:1988, p. 123-140) makes the same point. See also Brian Smith, *The Origin of Objects* (MIT Press, 1996).

The problem that I wish to highlight should not be apparent. The entire argument concerning the circular causal complexity of agent-environment interactions is vitiated, I believe, by its failure to engage the issue of strong representation. All the examples share (and must share) a certain problematic feature, viz, they are all cases in which the target behavior is *continuously driven* by the relevant environmental parameter. Yet a major motivation for the positing internal representations in the first place is to explain our puzzling capacity to go beyond tightly coupled agent-world interactions and to coordinate our activities and choices with the distal, the possible and the non-existent. The notion of internal representation is thus grounded in the notion of real inner surrogates and is merely extended (perhaps problematically) to the case of (merely) information-bearing inner states. This helps to explain why the best cases for the argument-from-continuous-reciprocal-causation may strike us as rather poor example of traditionally *cognitive* phenomena. For they depend crucially on the constant presence of the relevant environmental factors and thus do not strike us as especially “representation-hungry”²⁹ scenarios in the first place. Properly representation-hungry scenarios would be planning next year’s vacation, using mental imagery to count the number of windows in your old house, doing mental arithmetic, dreaming, etc., etc.

²⁹See Clark, A., & Toribio, J. (1994). Doing Without Representing? *Synthese*.

The dialectical situation is, however, rather delicate. For the anti-representationalist may now reply that the point of the argument, in part, is to suggest that these traditional cases (of what might be termed ‘environmentally de-coupled reason’) are in fact empirically marginal and that the bulk of intelligent response displays precisely the richly interactive profile the argument highlights. Environmentally de-coupled reason, it is claimed, is at best a ‘tip of the iceberg’ phenomenon. What is being promoted is thus a shift of emphasis away from off-line cogitation and onto real-time interactive engagement³⁰ -- a kind of ‘cognitive-to-coping’ shift.

This shift in emphasis is welcome. From both an evolutionary and a developmental³¹ point of view, real-world real-time responsiveness is clearly in some sense primary. But as part of any general anti-representationalist argument, the move is not dialectically suspect and empirically flawed. The problem is that the recognition that the richly interactive case is biologically basic is, as we shall see, perfectly comparable with the claim that ‘off-line’ environmentally de-coupled reason is not the mere tip of the adaptive iceberg. The way to forge a genuinely cognitive science of embodied, environmentally embedded agency is, I believe, to look harder for the bridges between densely coupled and strongly representationally mediated forms of adaptive success. Such bridges are at the heart of the conciliatory position I dub

³⁰This move is explicitly made in Haugeland, J. (1995). *Mind Embodied and Embedded*. In Y.-H. Houn & J.-C. Ho (Eds.), *Mind and Cognition* (pp. 3-38). Taipei, Taiwan: Academia Sinica and is also clearly in evidence in van Gelder, T., & Port, R. (1995). *It's About Time: An Overview of the Dynamical Approach to Cognition*. In R. Port & T. v. Gelder (Eds.), *Mind as Motion: Explorations in the Dynamics of Cognition* (pp. 1-44). Cambridge, MA: MIT Press.

³¹See Thelen, E., & Smith, L. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA: MIT Press.

‘minimal Cartesianism’, and to which I now turn.

4. MINIMAL CARTESIANISM.

Minimal Cartesianism seeks to locate the roots of pure contemplative reason in the richly interactive settings emphasized in the recent work on embodied cognition (Sections 2 and 3 above). Thus consider the phenomenon of skilled reaching³². Smooth, skilled reaching involves the use of proprioceptive feedback -- signals that tell the brain how the arm is oriented in space. But the timing of these signals poses a problem. The minimal delay between the onset and the use of such information looks to be between 200 and 500 milliseconds³³. Yet we make essential trajectory corrections, that look to be governed by such feedback, within the first 70 milliseconds³⁴ of reaching. How does nature turn the trick?

This problem of requiring feedback before it is practically available, crops up in industry too: in chemical plants, bioreactors and so forth³⁵. One common solution, in these cases, is to add a ‘forward model’ or ‘emulator’ into the systems. This is a circuit that takes as input a specification of both the previous state of the system and

³²I borrow this case from Grush, R. (1995). *Emulation & Cognition*. Ph.D. Dissertation, University of California. An extended treatment is available in Clark, A., & Grush, R. (1997). *Towards a Cognitive Robotics* (PNP Technical Report No.). Washington University in St. Louis.

³³This figure is established by, for example, using artificial vibrators strapped to the tendons to disrupt proprioceptive signal from the muscle spindles, and timing the gap between such disruptive input and alterations to the arm motion itself (see Dennier van der Gon, 1988, Redon et al., 1991).

³⁴See van der Meulen et al., 1990, Grush, R. (1995). *Emulation & Cognition*. Ph.D. Dissertation, University of California.

³⁵See Grush, R. (1995). *Emulation & Cognition*. Ph.D. Dissertation, University of California for a review.

the commands just issued, and that gives as output a *prediction* of the feedback that should later arrive. The emulator thus generates a kind of mock feedback signal available substantially in advance of the real thing.

Nature, it now seems, may deploy much of the same strategy. There is a fair body of neuroscientific evidence³⁶ that suggests that neural circuitry spanning the cortico-spinal tract, the red nucleus, the inferior olive, the contralateral thalamus and cerebellar cortex may be playing just such a role. Such circuitry looks to take a copy of the afferent motor command and to output a fast prediction of the feedback later due arrive by the slow 200-500 millisecond route.

The same trick has been replicated in a variety of neural network³⁷ models. What matters for our purposes, however, is an additional conjecture. It is the conjecture³⁸ that the biological emulator circuit plays a dual role. This dual role involves first the fine tuning of on-line reaching (the normal case, in which the emulator circuit acts as an aid to smooth real-time reaching). And second, the production of visuo-motor imagery allowing the off-line mental rehearsal of motor routines. In the later case, the emulator circuit is running alone, de-coupled from the real-world action system. Such an additional role for the very same emulator circuitry implicated in real reaching is

³⁶See Ito, M. (1984). *The Cerebellum & Neural Control*: Raven Press, Kawato, M., Furukawa, K., & Suzuki, R. (1987). A hierarchical neural network model for the control and learning of voluntary movement. *Biological Cybernetics*, 57, 169-185.

³⁷Kawato, M. (1990). Computational schemes and neural network models for formation and control of multijoint arm trajectory. In W. T. Miller III, R. Sutton, & P. Werbos (Eds.), *Neural Networks for Control*. Cambridge: MIT Press, Wolpert et al. (1995).

³⁸Grush, R. (1995). *Emulation & Cognition*. Ph.D. Dissertation, University of California.

independently plausible and helps explain some otherwise puzzling results. These include the robust finding that mental rehearsal can actually improve sports skills³⁹ and the activity of the cerebellum (generally thought of as a motor area) during mental imagery⁴⁰.

Motor emulation circuitry, if this is correct, is both an aid to fluent, real-world action and a support for independent, environmentally-decoupled mental rehearsal. It is a minimally Cartesian mental tool, but one that is parasitic upon adaptations closely geared to the promotion of real-time agent-environment interactions. As a result, the kinds of content that get represented are closely tied to the bio-mechanics and action-taking profile of the agent. And the form of the inner vehicles of content is left quite open -- such vehicles may involved complex temporally extended processes, as indicated in Section 2.

Given such a profile, we can see why isolationist methodologies and assumptions are inadequate even in the case of certain kinds of environmentally decoupled cognitive skills. For such skills may remain action-oriented at one x with

³⁹Fetz & Landers (1983).

⁴⁰Decety et al. (1990).

both contents and mechanisms being profoundly informal by the agents real-time interactive purposes. Once again, however, the failure of isolationism should not be seen as an invitation to scepticism about representation and inner models. In the emulator case, it is clearly apparent that we are dealing with identifiable circuitry whose functional role is to model specific aspects of extra-neural (in this case bodily) reality. Yet this inner modeling is of a type that is perfectly continuous with the various morals and emphasizes suggested by the action-oriented research discussed earlier. The conciliatory position that I favor this involves combining the stress on real-world, real-time action with a search for the biologically basic roots of more decoupled forms of thought and problem-solving. It is only by confronting the latter class of cases that representationalism can be given a fair trial.

5. SCALING, RATIONALITY AND COMPLEXITY.

Minimal Cartesianism aims to build bridges between the recent emphasis of richly interactive tasks and the more traditionally cognitive focus on decoupled reason. To that end it stresses the use of multiple, partial, action-oriented inner models and of deictic, idiosyncratic and action-oriented internal representations. The compelling question, at this point becomes whether we must counteract, *in addition* to these, inner resources that have much of the look of classical symbolic models. Can we hope to explain the full gamut of human cognition without at some point reinventing the original image of context-neutral, rich, action-independent, highly manipulable inner symbolic structures?

It is, I think, worth a try! More accurately, what may be worth a try is an approach which does not eschew altogether the use of such richly structures, action-neutral

encodings but which ties them very closely to our experiences with public language and other externalizable and interpersonally shareable symbol systems. Complex human cognition would then be depicted as the fecund interface between a variety of action-oriented internal resources and a larger web of linguistic competence and cultural tools and practices.

This larger web (or ‘scaffolding’) acts so as to substantially alter the computational spaces that can be explored by our form of basic, on-board biological reason. A classic example⁴¹ is the use of pen and paper to expand our basic mathematical horizons, allowing us to use an iterated sequence of simple inner computations (7 x 7, 4 x 4) to solve more complex problems (such as 777 x 444). Public language, I elsewhere argue plays a wide variety of similar roles⁴². The mere act of labeling, as Dennett⁴³ points out, affords great economies of search and classification. While the capacity for linguistic rehearsal may, according to Ray Jackendoff⁴⁴, enable us to attend to the details of our own thoughts and thus open up new possibilities of reflection and analysis⁴⁵. External artifacts and social organizations

⁴¹Rumelhart, D., & McClelland, J. (1986). On learning the past tenses of English verbs. In D. Rumelhart & e. al. (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Vol. 2, pp. 216-271). Cambridge: MIT Press.

⁴²Clark (in press), Magical words: How language augments human cognition, in P. Carruthers (ed) *Thought & Language*.

⁴³Dennett, D. (1995). *Darwin's Dangerous Idea*. New York: Simon & Schuster.

⁴⁴Jackendoff, R., How language helps us think, *Pragmatics & Cognition* 4:1:1996, p. 1-34.

⁴⁵Ibid, p. 19-22.

likewise alter and transform the tasks that individual brains need to perform. And Ed Hutchins⁴⁶ offers a persuasive account of ship navigation in which it is the overall system comprising multiple brains, bodies and instruments that ‘solves’ the problem. Each crew member within this larger nexus, merely monitors and responds to certain conditions, and alters a few aspects of the shared work space so as to support the activity of the others. The whole process constitutes an environmentally extended computational flow in which props and artifacts (such as nautical slide rules) also play a major role.

The minimal Cartesian treatment of basic biological reason may (just may) thus scale-up so as to illuminate the full panoply of human thought and reason. But it will only do so if we take the issue of external scaffolding very seriously and recognize the special virtues of public language: the one action-neutral symbolic code we know ourselves and possess. One implication of this approach to the scaling problem is that we will need, at times, to study these larger systems (of multiple communicating brains and artifacts) as organized wholes and to recognize extended computational processes spanning the boundaries between brain, body and world. Such assertions can easily be mistaken for antipathy towards the study of the inner resources and processes. But the real challenge, once again, is to interlock the two approaches and

⁴⁶Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA: MIT Press.

thus to relocate individual human reason in its proper ecological niche.

All this raises questions about the notion of human rationality itself. Isolationist cognitive science tended to depict rationality in terms of semantically apt transitions between inner mental states. Turing's achievement, as repeatedly stressed by Jerry Fodor⁴⁷ was to show how such transitions could be supported by a mechanical process. The environmentally extended approach just mooted does not reject that account. It may (and should) incorporate Turing's central idea of inner processes whose syntactic⁴⁸ properties preserve semantic relations. But this will be just part of a larger theory that allows rational behavior to supervene on wider webs of structure involving other agents and aspects of the local environment.

⁴⁷For example, see the comments on p. 277-278 of his *Replies to Critics*, in B. Loewer & G. Rey (eds) *Meaning in Mind: Fodor & his critics* (Blackwell: Oxford), p. 255-319.

⁴⁸Syntactic properties are any non-semantic properties that can be directly exploited by a physical system. Temporally extended processes, as described in Section 2, are in this sense syntactic too.

Finally, a worry about complexity. Even if the general project sketched in this paper proves attractive (the project of bridging between interaction-based models and environmentally decoupled, representation-involving, reason), there remains a worry concerning the potential complexity of the inner vehicles of content. The worry already touched on in Section 2, is that the inner vehicles may be too spatially and temporally complex to effectively isolate. The worry gains force from recent demonstrations of the role of recurrent connections⁴⁹ in modulating the information-processing profile of neuronal populations⁵⁰ and from the sheer difficulty of assigning specific content bearing roles to tracts of neural machinery. These complexities and difficulties can lead to a subtly different kind of scepticism in which it is the complexity of the inner story itself (not the inner-outer interaction) that is supposed to make trouble for the representational analysis.

The issues here are more purely empirical and it is impossible given the current state of research, to make any firm predictions. But one clear possibility is that new analytic tools may yet provide the means to identify functionally important patterns of activity. Dynamical systems analyses, of the kind sometimes promoted as an alternative to the representational approach, may in fact serve against the backdrop of burgeoning spatial and temporal complexity. This possibility is clearly noted by van Gelder⁵¹ himself, who allows that “an exciting feature of the dynamical approach is

⁴⁹Knierim, J., & VanEssen, D. (1992). Visual Cortex: Cartography, Connectivity and Concurrent Processing. *Current Opinion in Neurobiology*, 2, 150-155.

⁵⁰“What might be cognition be...” op cit, p. 376.

⁵¹Of course there must be something that persists or else memory-based action would be

that it offers opportunities for dramatically reconceiving the nature of representation in cognitive systems.” Internal representations, then, may be realized not as simple inner states but as dynamical patterns of just about any conceivable kind. Such patterns may, in addition, be transient entities that form only in response to the details of current context. We thus better appreciate the limits of the “inner vehicle” metaphor itself. Such vehicles need be neither simple nor persisting⁵² in order to play a representational role .

Van Gelder’s concession is important. He does not take himself to have shown that there are no internal representations. Just that there might not be any, and that if there are they may take a very different form to the one we once expected. I have tried to show that the specific skeptical considerations he advances (concerning the potential complexity of agent-environment interactions) fail (and must fail) to make

impossible. The point is just that the space of internal representational vehicles may be much larger than the space of persisting inner states.

⁵²See papers in Port, R., & Gelder, T. V. (Eds.). (1995). *Mind as Motion: Dynamics, Behavior, and Cognition* . Cambridge, MA: MIT Press.

contact with the representationalist hypothesis, which is grounded in our capacities for environmentally decoupled reason. The ‘revisionary representationalist’ option, however, is both appealing and increasingly in evidence in actual cognitive scientific applications.

In sum, our vision of basic biological reason is rapidly changing. There is a growing emphasis on the computational economies afforded by real-world action and our growing appreciation of the way larger structures (of agent and artifacts) both scaffold and transform the shape of individual reason. These twin forces converge on a rather more minimalist account of individual cognitive processing -- an account that tends to eschew rich, all-purpose internal models and sentential forms of internal representations. Such minimalism, however, has its limits. Despite some ambitious arguments, there is currently no reason to doubt the guiding vision of individual agents as loci of internal representations and users of a variety of inner models. Rather than opposing representationalism against interactive dynamics, we should be embracing a broader vision of the inner representational vehicles and seeking the continuities between rich interactive strategies and off-line, environmentally de-coupled, reason. The reward may be a better vision of reflective agency itself.