

The Dynamic Emergence of Representation

Mark H. Bickhard

A final version of this paper is in press as:

Bickhard, M. H. (in press). The Dynamic Emergence of Representation. In H. Clapin, P. Staines, P. Slezak (Eds.) *Representation in Mind: New Approaches to Mental Representation*. Praeger.

Mark H. Bickhard
Department of Philosophy
15 University Drive
Lehigh University
Bethlehem, PA 18015
610-758-3633 office
610-758-3775 philosophy dept
mhb0@lehigh.edu
mark.bickhard@lehigh.edu
<http://www.lehigh.edu/~mhb0/mhb0.html>

The Dynamic Emergence of Representation

Mark H. Bickhard

Representation did not exist moments after the Big Bang; it does now. Representation has emerged. Accounting for that emergence is among the central problems of naturalism today. The emergence of representation is a difficult problem for a number of reasons, but one of them is that representation involves issues both of fact and of normativity, and accounting for the naturalistic emergence of normativity is particularly difficult — some would say impossible.

The possibility of ontological emergence of any kind is itself in question, according to some arguments, so I begin by examining one of the most powerful of those arguments and deriving a metaphysical lesson from it: accounting for non-epiphenomenal ontological emergence requires a process metaphysics. Next, a basic flaw in the argument against the possibility of deriving normativity within a naturalistic framework is pointed out, thus freeing the discussion to develop a naturalistic model of the emergence of representation.

Emergence

Jaegwon Kim has presented an argument against the possibility of genuine emergence that, if sound, would render nugatory all attempts to account for any kind of emergence (Kim, 1989, 1990, 1991, 1992, 1993a, 1993b). I contend that the argument is not correct, but that a deep lesson can be derived from it.

Kim argues that genuine emergence must involve the emergence of causal power. The emergence of mentality, for example, if it were causally inert, would at best be a form of epiphenomenalism. But this creates a dilemma if we are to hold to basic naturalistic commitments: 1) Either all causal power is resident in whatever the lowest level of particles may be, in which case all higher level processes (including mental processes) are causally epiphenomenal, or 2) some new causal power does come into existence, in which case the world is not physically closed, and naturalism, at least in most conceptions of naturalism, is false. So, the dilemma is between the failure of emergence or the failure of naturalism.

The first horn of the dilemma is simply that no high level entity or process, including mental processes, can be anything more than the working out of the causal interactions of the basic particles involved. Any new causal regularities that might be observable are merely regularities of that particular kind or organization of particle interaction; all causal power remains at the level of the particles per se, and there is no emergent causal power.

If there *is* presumed to be some new causal power, then that causal power will be ad hoc relative to particle level causality, and will entail that the physical, the particle, universe is not causally closed. This introduces a kind of causal dualism (or worse) between particle causality and whatever new causality comes into existence and constitutes, among other things, a failure of naturalism. Thus the second horn of the dilemma.

The argument, however, though valid, is unsound. The crucial false premise is a metaphysical premise, and Kim's argument, thus, can be taken as a *reductio* of this premise. The false premise is the assumption of a particle metaphysics.

Put simply, there are no particles. The best contemporary physics tells us that the world is composed of quantum fields, not particles (Brown & Harré, 1988; Davies, 1984; Saunders & Brown, 1991; Weinberg, 1977, 1995, 1996). Field models are mathematically forced by special relativity together with conservation of energy. What appear to be particles are the quantization of quantum wave phenomena; such quantized

phenomena are no more particles than is the “quantization” of the number of waves in a guitar string. That is, particles are the “appearance” of quantized quantum field phenomena, and have no substantial existence beyond that. Furthermore, we know of phenomena that are simply inconsistent with a particle based physics, but are predicted by quantum field theory, such as the Casimir effect (Aitchison, 1985; Sciama, 1991; Weinberg, 1995). Quantum field theory, therefore, provides its own empirical predictions and support. Still further, there are in-principle arguments against a strict particle physics, and, therefore, against a strict particle metaphysics: e.g., point particles would have zero probability of ever encountering each other, and, therefore, nothing would ever happen in such a universe (Bickhard, 2000).

A standard semi-sophisticated contemporary view might postulate both particles and fields, with the fields enabling the interactions among the particles. This would be incorrect, but what I argue is the crucial point is already accepted even in this move: fields are processes, and the conclusions of Kim’s argument require that we move to a process metaphysics.

The reason for this rests on a diagnosis of Kim’s basic argument. In framing the anti-emergentist argument within a particle metaphysics, Kim has presupposed that causal power is resident in particles. Particles do not themselves have any organization. They are points. Particles, however, can participate in organization in space and time and with respect to each other. It is precisely within such organization that particles engage in their presumed causal interactions. Such organization is the stage setting within which particle interactions proceed.

The critical move here is that organization is relegated to a strictly stage setting role. Causality is possessed by entities that do not have organization, and organization itself is not a legitimate locus of causal power: it is *merely* stage setting. It is this delegitimation of organization as a potential locus of causal power, and, therefore, as a potential locus of *emergent* causal power, that drives Kim’s argument.

Processes, in contrast, are inherently organized. A point process is a nonsensical notion. If everything is process — and if everything is quantum field theory, then everything *is* process — then everything that has causal power is itself already organized. Organization cannot be delegitimated as a locus of causal power without emptying the universe of all causality. Furthermore, causal power varies with organization; some process organizations have different causal powers than other organizations. So, a new organization of process might well have a new causal power, by virtue of that organization. Such a new causal power will be emergent in that organization. This is a naturalistic and non-epiphenomenal emergence.

In sum, then, Kim’s argument shows that emergence makes sense only within a process metaphysics, and the clear fact of emergence therefore forces a process metaphysics. It is doubly fortunate, then, that contemporary physics also forces a process metaphysics.

Normative Emergence

It is a standard truism that “ought” cannot be derived from “is” — norms cannot be derived from facts. But it is a false truism. Because I propose precisely to show how certain norms relevant to representation can be emergent from, and therefore derived from, facts, I will clear the way for that by showing how the argument against deriving norms from facts fails.

The core of the argument is the claim that normative terms cannot be validly introduced in a derivation from premises that do not already contain normative terms.

Since facts, presumably, do not contain normative terms, then no normative conclusions can be derived.

There is actually more than one point at which this argument can be attacked, but I will focus on one in particular. The structure of the argument assumes that the only valid form of introduction of terms is via abbreviatory definition. If every valid new term is simply an abbreviation for combinations involving prior terms, and if the base set of terms contain none that are normative, then any valid conclusion could be rendered solely in the original non-normative terms by back-substituting through the definitions. By assumption, then, the conclusion would not involve norms.

But the assumption that the only valid form of definition is abbreviatory definition is false. There is also implicit definition. The model theoretic notion of implicit definition is illustrated by the sense in which a set of formal sentences implicitly defines the class of models that would satisfy those sentences. Hilbert argued, for example, that the axioms of geometry implicitly defined the terms of geometry, such as “point” and “line”. Implicitly defined terms cannot be substituted away in favor of some base set of terms. This point does not demonstrate *how* to derive norms from facts, but it does show that a commonly presumed impossibility of doing so does not hold.

There is, however, a rejoinder to this point that I would like to address. Beth’s theorem (in model theory) demonstrates that implicit and explicit definition are of equal power (Chang & Keisler, 1990), and this is often taken as a reason for ignoring implicit definition (Doyle, 1985). This stance, however, overlooks important points. First, the equivalence between implicit and explicit definition is extensional only. Second, and more deeply, Beth’s theorem holds in the case of first-order predicate logic with infinite models. This is not the only combination of logic and model theoretic framework. In every combination examined of logics, e.g., first order predicate logic, infinitary logics, fixed-point logics, and models, e.g., infinite models, finite models, implicit definition is either of equal power to explicit definition, or it is more powerful than explicit definition (Dawar, Hella, Kolaitis, 1995; Hella, Kolaitis, Luosto, 1994; Kolaitis, 1990, manuscript 1996). Implicit definition, therefore, cannot be ignored, and it suffices to render “no ‘ought’ from ‘is’” invalid.

Representation

The model of representation that will be outlined in the following pages is called interactivism. Interactivism models representation as emergent in a particular kind of biological function, so the first focus is to model the emergence of biological function. The normativity of representation derives from that of biological normative function (Bickhard, 1993).

In this structure, the interactive model resembles etiological models, such as that of Ruth Millikan (Millikan, 1984, 1993), which also first model biological function and then representation as a special case. The specifics, however, differ fundamentally: 1) the interactive model of function is not etiological, and 2) the relationship between representation and function is not the same. Before proceeding to the interactive model, I will show that etiological models are not acceptable as models of the naturalistic emergence of function and representation. In particular, they are causally epiphenomenal, and, in that sense, fail naturalism (Bickhard, 1993).

Etiological Models of Function and Representation. In an etiological model of function, X having a (proper) function is constituted in X being the product of the requisite history of evolutionary selections. A kidney, for example, has the function of filtering blood because the evolutionary ancestors of this kidney were selected for filtering blood. Conversely, it is because kidneys have the function of filtering blood, and that the

ancestors of this kidney succeeded in fulfilling that function, that this kidney exists. There are many details to attend to and sophisticated and ingenious elaborations of and derivations from these notions in the literature, but the crucial point that I wish to focus on is that function is *constituted* in having the right history. To have a function *is* to have the right history.

This point has some counter-intuitive consequences; the consequence that I wish to focus on is one which Millikan has partially addressed with the following science fiction example. If a lion were to suddenly pop into existence in the room that was, by hypothesis, molecule for molecule identical to the lion in the zoo, nevertheless, the heart of the lion in the zoo would have a function, while none of the organs in the magical lion would have any functions at all. No part or parts of the lion that just appeared have any history at all, and certainly, therefore, no evolutionary history. Correspondingly, they do not have the right evolutionary history to constitute them as having any functions. There would be no biological normativity involved at all in the suddenly-appearing lion, despite its being molecule for molecule identical to the zoo lion.

This is strongly counter-intuitive, but quantum mechanics has taught us that counter-intuitiveness is not necessarily a good ground for rejecting a theory. It might seem, in fact, that such a consequence is a small price to pay for the power of a naturalized model of biological function, and rendering function in terms of biological evolution certainly has the *prima facie* look of a successful naturalization. This seems to be Millikan's stance.

The example, however, illustrates an even deeper problem. Etiological function is ontologically constituted in having the right history. Having the right history is not a property of the current state of a system. The two lions, for example, have identical current physical states by hypothesis, yet they do not have the same functional state because they do not have the same histories. But only current state can be causally efficacious. The two lions will have identical causal properties, but one will have organs with functions and the other will not. That is, etiological function is causally epiphenomenal; it makes no causal difference in the world. Etiological function is not a naturalization of a causally relevant function.

I turn, therefore, to a model of the emergence of function that I claim *is* naturalistic, including accounting for emergent causal power.

Autonomy, Self-Maintenance, and Function. In a substance or particle metaphysics, the stability through time of the basic substances or particles is taken for granted. In a process framework, any stability of an organization of process requires explanation. There are two general forms of such explanation: 1) Some organizations of process are in energy wells, in the sense that a change in the organization would require the introduction of energy above what is currently impinging on the process. Atoms, molecules, and much of the standard furniture of the world is temporally persisting because of such energy well stabilities. 2) The second form of such stability is that of processes that are far from thermodynamic equilibrium. Such a process will move toward equilibrium, and thus cease to exist, unless some active counterinfluence is operative. That is, they are unstable if isolated from such counter-equilibrating influences.

In some cases, those influences are completely external to the system itself. A chemical bath can be maintained in a far from equilibrium condition, for example, with the pumping into the chamber of appropriate chemicals. Any stability is dependent on the continuing operation of the pumps and availability of the chemicals. Such systems can exhibit interesting properties, such as self-organization. They also illustrate a fundamental fact about stable far from equilibrium systems: they are open to their environments, not just as a matter of fact, but as a matter of ontological necessity. If cut off from their environments, they cannot remain far from equilibrium, and they cease.

Self-Maintenant Systems. A more interesting case for current purposes, however, is the class of far from equilibrium systems that make their own contributions to their own stability. A canonical example is a candle flame. A candle flame maintains above combustions threshold temperatures; it vaporizes wax into flammable gases; and in standard atmospheric and gravitational conditions it induces convection, which pulls in fresh oxygen and gets rid of waste products. A candle flame is, in several ways, *self-maintenant*.

Recursive Self-Maintenance. A self-maintenant system can maintain itself over some range of conditions — if a candle is put into a vacuum or doused with water, it ceases. Some systems, however, can, in addition, contribute to their own stability over a range of *changes* in conditions. They can change what they do to maintain stability in accordance with changes in environmental conditions. A bacterium, for example, might swim and continue swimming if it is going up a sugar gradient, but tumble if it finds itself swimming down a sugar gradient (D. T. Campbell, 1990). It maintains its condition of being self-maintenant in the face of changing environmental conditions: it is *recursively self-maintenant* (Bickhard, 1993).

Recursive self-maintenance requires some way of differentiating environmental conditions, two or more differing ways of contributing to conditions for stability, and appropriate switching relationships between the differentiations and the alternative potential kinds of interaction. Such properties require that the system have infra-structure: process organizations that are on a sufficiently slower time scale that they can be taken as fixed, as structural, during the differentiations, switching, and interactions. E.g., processes that are on a sufficiently slower time scale than those of the swimming and tumbling that they can control the process of switching between swimming and tumbling. The (self) maintenance of such infrastructure is what constitutes metabolism (Moreno & Ruiz-Mirazo, 1999).

In the case of a recursively self-maintenant system, the infrastructural processes have both an informational aspect — differentiation and switching — and an energetic aspect — guiding energy into swimming. This suggests a more minimal form of infrastructure. Some systems may have infrastructure that accomplishes work — that guides energy — but that has no informational aspects. An example might be a sulfur based cell that does only one thing, “burn” sulfur, but does it with complex metabolic processes.

The models of self-maintenance, work infrastructure, and recursive self-maintenance collectively constitute a graded model of autonomy (Bickhard, in press). Autonomy is the ability of a system to maintain itself in accommodation to its environment, and self-maintenance, infrastructure, and recursive self-maintenance constitute increasing grades of such autonomy.

There is now in place a sufficient model to address both function and representation. Function first: Serving a function is modeled as making a contribution to far from equilibrium stability. Serving a function, therefore, is relative to the system which is being contributed to. A heart, for example, may serve a function for a parasite, but be dysfunctional for the host. The normativity of function will be similarly contextualized. Note that serving a function contributes to the stability of a far from equilibrium process, which has distinct causal consequences in the world: this is not a model of an epiphenomenal function.

It should be noted that this explication turns upside down the explicatory organization in etiological models. Etiological models take as primary the notion of *having* a (proper) function, and derive a notion of *servicing* a function from it: X *servicing* a function is (roughly) X accomplishing a function that X has.¹ Here, the notion of *servicing* a function is primary. Basing a model of *having* a function on this model of *servicing* a function requires one additional preliminary notion.

Dynamic Presupposition. A recursively self-maintenant system may just switch from one interaction with its environment to another as differentiated conditions change, such as is the case for the swimming and tumbling of the bacterium, or it may set up indications of multiple interactions that would be appropriate in current circumstances, and engage in some more complicated process of (inter)action selection. That is, action selection can occur via simple triggering, or via more complex selection processes among indicated interaction potentialities.

There is much to be addressed about such systems of action selection, but the crucial point for now is that any triggering of an interaction, or any indication of the current appropriateness of an interaction, presupposes that that interaction is in fact appropriate for the current conditions. Continuing to swim *down* a sugar gradient is, in general, *not* appropriate. Appropriateness here is a normative notion, and the normativity is a functional normativity. That is, it is derived from the norm of contributing to the maintenance of the conditions for the far from equilibrium continued existence of the system. Interaction (types) will tend to be appropriate in some conditions, and not in others. An indication of the appropriateness of an interaction, therefore, *dynamically presupposes* that those conditions obtain. The dynamic presuppositions of an interaction or interaction indication are those conditions that would make that interaction appropriate, that render it likely to make a functional contribution. More generally, a process dynamically presupposes whatever those conditions are, *internal to the system or external to the system*, that support its being functional for the system.

Having a function. *X has a function* insofar as some function being served dynamically presupposes one or more consequences of X. X has a *proper* function insofar as it is dynamically presupposed by embryological or metabolic functional processes — that is, it has a proper function if it is dynamically presupposed in the biological “design” of the system. A kidney has a (proper) function of filtering blood because the organization of the other organ systems in the body dynamically presuppose that something at that relative location will filter blood.

Representational content. The dynamic presuppositions of a blood circulatory system will in general be internal: hearts and kidneys, for example. The dynamic presuppositions of an interaction indication will be about the environment. If those dynamic presuppositions do not hold, then the interaction will fail. That is, if those dynamic presuppositions are *false*, the interaction will fail. Dynamic presuppositions, then, can be true or false, and they can be true or false about the environment. Dynamic presuppositions constitute representational content about the environment.

Interaction indications, in this model, are the primitive form of representation. They predicate of the environment that the environment possesses the dynamically presupposed conditions. They predicate that *content* of the environment.

Such an interactive representation may be false: the dynamically presupposed conditions may not be true. Furthermore, they may be (fallibly) discovered to be false: if the system engages in the indicated interaction, and it does not proceed as indicated, then the dynamic presuppositions, the content, was false. In this model, not only the possibility of error, but also of system detectable error, are trivially accounted for.

This is indeed a primitive form of representation. More needs to be addressed to indicate its potential to be ground for all representation. It is also a model of representation that has several unfamiliar properties — properties not common in standard models. These too will be outlined.

More Familiar Representations: Objects. I will address first how the interactive model could account for the representation of physical objects. If an organism differentiates a relevant condition in its environment, it will invoke indications of

appropriate further interactive potentialities. Even when that differentiation process is inactive, however, the control infrastructure that would engage in it, and its relationships to interaction indications, are still present in the system. Such an aspect of the control structure constitutes a conditionalized indication of interaction potentialities: if XYZ differentiation is made, then QRS interactions will be indicated. Conditionalization, in turn, creates the possibility of iterating such indications: if XYZ differentiation occurs, then QRS is possible, and if QRS occurs, then ABC will be possible.²

So, interaction indications can both be multiple — they can branch — and they can iterate. As such, they can form webs of interconnected conditionalized indications of interaction potentiality — perhaps vast and complex webs. Some subwebs of such a larger web may come to have special properties. In particular, they may be internally reachable, in the sense that any indicated interaction anywhere in the subweb is reachable as a direct interaction potentiality, perhaps via various intermediary conditional interactions, and that internal reachability property may remain invariant under some relevant class of other kinds of interactions. For example, a child's toy block will afford multiple potentialities of visual scans and manipulations. Any one of these potentialities is available from any other — e.g., you can always turn the block back so that an earlier visual scan is again possible — so the subweb of interactive potentialities for this block is internally reachable. Any that internal reachability itself remains invariant under a large class of other interactions, such as putting the toy away in the toy box, the child leaving the room, and so on. It is not invariant under all possible interactions, however, such as crushing or burning the block.

This outlines the general manner in which the interactive model can scale from simple interaction possibility representations to representations of physical objects. It is a generally Piagetian, or pragmatic,³ model of object representation, and I would suggest a generally Piagetian approach to other more complex kinds of representation, such as abstract representations — such as of numbers.⁴

What about Input Processing? Models of representation are standardly what the pragmatists called spectator models. They are models of some homunculus staring back down the input stream, processing inputs, rather than future oriented models of interactive anticipation. But input processing clearly does occur — in sensory systems, for example. If such input processing is not to be taken as somehow constituting or generating representation, what account is to be given of it?

The interactive model distinguishes between two aspects of epistemic relationship to the world: contact and content. Contact with the environment is provided by the differentiations of that environment. Such differentiations are the basis for setting up indications of further interactive potentialities; they are how the system can locate itself in its web of conditional interactive indications. Without contact, no interactive content, no indications of potentiality, would have any likelihood of being appropriate for any particular environment.

Such indications, in turn, constitute representational content. It is such anticipatory indications that involve dynamic presuppositions, presuppositions that can be false. It is in these presuppositions that representation is emergent.

Differentiation in general is generated by the internal outcomes of previous interactions. If an interaction control system is engaged in interaction with an environment, the internal course of that interaction will be partially determined by the control system, but importantly determined by the environment. Differing environments will yield differing internal courses of the interaction, and differing internal outcomes of the interaction. Any particular possible outcome of an interaction serves to differentiate those environments that would yield that outcome from those that would yield a different outcome: the outcomes differentiate types of environments. There is no other information available in such a differentiating outcome per se about what kind of environment it differentiates, but

nevertheless it may be useful for setting up indications of further interactive potentialities. If so, then any such indication predicates of that environment whatever properties are dynamically presupposed by those indications. It is the future oriented indications that represent (something about) the differentiated environment, not the differentiations per se.

Differentiations in general may involve full interactions, but a simple version would be a differentiation process that had no outputs, a passive differentiation. A passive differentiation is a differentiation nevertheless, and can serve as the basis for further indications of interactive potentiality.

But passive differentiations are just input processing. Input processing, then, is an aspect of the interactive model just as it is for spectator models. The difference is that standard models take input processing as constituting or generating representation, while the interactive model takes it to be only a simple case of the general function of differentiation — of contact. In effect, input processing models conflate contact and content; they take whatever the contact is in fact with as somehow the content of the purported representation.

Properties of Representation

It is a large programmatic task to demonstrate the adequacy of the interactive model for all forms of (purported) representation — perception, memory, rational thought, language, and so on. These have been addressed elsewhere.⁵ For current purposes, I will take it as demonstrated that the interactive model is a candidate for a model of the nature of representation, and proceed to examine some of the consequences of that nature.

Representational Error. As pointed out earlier, the possibility of representational error is trivially accounted for in the interactive model: the dynamic presuppositions may be false. This is in contrast to correspondence models of representation, that simply do not inherently have the resources to account for error, and must, at best, superimpose some additional criterion for error on the basic correspondence framework. The limitation is, simply, that if a purported representation constituting correspondence exists, then the representation exists and is correct, while if the crucial correspondence does not exist, then the representation does not exist. There are only two model possibilities — the correspondence exists or the correspondence does not exist — but there are three conditions to modeled — the representation exists and is correct, the representation exists and is false, and the representation does not exist.⁶

One attempt to introduce such an error criterion is Fodor's asymmetric dependency criterion. Consider two conditions under which a representation is invoked, one purportedly correct and the other incorrect. If the representation is constituted simply in the invocation relationship (be it causal, nomological, informational, or whatever), then the purportedly incorrect deployment of the representation is just as legitimate a participant in the representational constitution as is the "correct" object. So, if the objects are X and Y, there are no grounds for the claim that the representation is supposed to represent Xs and that its invocation for Y is in error. Instead, since both Xs and Ys activate the representation, the content should be construed as "X or Y" and the possibility of error evaporates. This so-called "disjunction problem" is just one version of the general problem of accounting for representational error.

Fodor has suggested that the correct and incorrect cases can be distinguished in the following way: the incorrect invocation is dependent on the correct invocation in the sense that the incorrect deployment would never occur if the correct case didn't exist, but that dependency is not reciprocated; it is asymmetric in the sense that the correct case could occur even if the incorrect case never did. In the by now canonical example, if the COW representation is invoked by a horse on a dark night, that is in error because such "horse on

dark night” invocations are asymmetrically dependent on invocations by cows (Fodor, 1990, 1991)

There are multiple problems with Fodor’s model, but a straightforward counterexample to the asymmetric dependency criterion is the following: Consider a neurotransmitter docking on a receptor in a cell surface and evoking corresponding activities in the cell. Here we have full biological and nomological correspondences. Now consider a poison molecule that mimics the neurotransmitter and also docks on that receptor. Here is a clear case of asymmetric dependency, yet at best we have a case of functional error, not representational error. Fodor cannot account for the possibility of representational error (Bickhard, 1993; Levine & Bickhard, 1999).

Millikan, because of the distinction between what something is functionally supposed to represent and what it is being used to represent currently, can model the possibility of representational error. But Millikan’s model is epiphenomenal, and thus not acceptable for a different reason. Similarly, Dretske (1988) has resources for modeling representational error, but this model too is etiological (a learning etiology), thus epiphenomenal (Bickhard, 1999).

Cummins invokes a distinction between the target to be represented and the representation being applied to that target (Cummins, 1996). In terms of this distinction, he models error as a representational content being applied to a target to which it is false. This is, I would argue, roughly the correct strategy for accounting for representational error. Cummins’ distinction between target and representation is capturing the distinction in the interactive model between a differentiated condition and the interactive content attributed to it, between contact and content (Bickhard, 1980, 1993). I argue, however, (see below) that Cummins’ model fails to satisfy other crucial criteria.

System Detectable Error. In the interactive model, if an indicated interaction is undertaken and the interaction does not proceed as indicated, then the indication is false, and is falsified for the system itself in a way that is potentially usable by that system. Representational error is system detectable. Only if error is system detectable can it be used to guide further behavior or to guide learning processes. Clearly system detectable error occurs, and, therefore, any model which cannot account for it is impeached.

In general, models of representation do not even address the criterion of system detectable error. It is clear, however, that standard models cannot account for it (Bickhard, 1999). No organism can take into account the evolutionary or learning history of its functional representations, or the asymmetric dependencies among potential invocations of its representations, to be able to determine what its representations are supposed to represent. Nor can they then compare that normative content with what is currently being represented to find out if the representation is being truly applied or falsely applied — to accomplish the later is the problem of representation all over again.⁷

Cummins’ Model. In the case of Cummins’ model (Cummins, 1986), the normativity of the representations is carried in the observer of the system, not in the system itself. This is illustrated by one of Cummins’ examples: Consider a toy car that is to run a maze. The car’s wheels are guided by a peg that slides back and forth in a slot in a card that is inserted into the car and that is read by moving the card progressively through the location of the peg. Differing cards with differing slot patterns would, thus, guide the car through differing mazes. The slot in a card is isomorphic to the required turns in the mazes that that card correctly represents. This is an instance of Cummins’ general point that representational content is constituted in structures that are isomorphic to what they are to represent.

The normativity in this example, however, is carried in the assumption that the car is supposed to run through the maze, and that assumption is made by the observer, not by

the car. If the goal were to hit the wall of the maze at a certain point, then the card that gets the car through the maze would no longer be correct.

This example also helps to illustrate another point. Cummins assumes that the notion of structural isomorphism is relatively unproblematic, but it is in fact seriously problematic. There is no “fact of the matter” about what the relevant structure is in any material entity. Suppose that the card inserted into the car were not “read” by a peg in the slot, but, rather, by a head responding to the pattern of magnetic domains along the edge of the slot. Now the slot pattern per se is irrelevant.

Structure, then, must be relative to the norm of the interactive task that the system is engaged in. That norm, in turn, must be a norm emergent in the system itself. And structure is relative to the method of “reading”, which is a control system notion, not a material notion. Further, the only possible consequence that such a “structural” read out could have is to influence the course of the (interactive) processes in the system, and any structure per se is in-principle superfluous for accomplishing such process flow influence (Bickhard, 1980). The crucial property is to control the interaction in such ways that it proceeds as indicated, as selected for, in the differentiated environments — and that is the interactive model.

Future Orientation. Correspondence models of representation are past oriented, with the input processing spectator looking backwards down the input stream. The interactive model is future oriented. Representation is, most fundamentally, of future potentialities of interaction. Future orientation is a feature of pragmatist models generally, but is rarely found in contemporary models. It is the future orientation of the interactive model that makes accounting for error and for system detectable error so immediate.

Modality. Interactive representation is of future potentialities of interaction — that is, representation is of *possibilities*. Interactive representation, then, is inherently modal. Standard models rarely address this issue, but the presumption is that representation is of actualities (whatever is actually on the other end of the input stream) and that modality is something to be added or dealt with later. Interestingly, young childrens’ cognition is inherently modal, with actuality, possibility, and necessity being poorly differentiated, rather than being amodal with modality developing later (Bickhard, 1988; Piaget, 1987).

Implicitness. Interactive content is the dynamic presuppositions made in indications of interactive potentiality. Those presuppositions are not explicitly represented; instead, they are implicit in the indications themselves. It can be explicit that an interaction of a particular kind, arriving at a designated outcome, indicates that one or more further interactions would be possible, but what supports those indications, what is presupposed about the environment by those indications, is not explicit.

This implicitness of content is fundamentally different from standard models. Encodings cannot be encodings without explicit content. Implicitness is a source of some of the power of the interactive model — for example, I argue elsewhere that the frame problems arise largely from attempting to render implicit content in explicit form (Bickhard & Terveen, 1995).

The interactive model easily accounts for the possibility of representational error, as well as the possibility of an even stronger criterion: system detectable representational error. It also has the consequences that representation is future oriented, modal, and, at base, implicit. In all these respects, it differs radically from standard models.

Three Desiderata

This discussion introduces three criteria that I would like to propose as requirements for an acceptable model of representation. These three are not exhaustive, nor

are they strictly independent, but they do, arguably, capture necessary aspects or facets of representation, and I find them useful to illustrate problems with models in the contemporary literature. The three are:

- 1) Representational content must be internally related to whatever is to constitute representation.
- 2) Representational content must be functionally accessible to the system for which representation is to be modeled. And
- 3) Representational content must be normative for the system for which representation is to be modeled.

For short, the relationship between representation and content must be system internal, system accessible, and system normative. Only the first of these requires any significant discussion, and that is because the distinction between internal and external relations is no longer widely understood or appreciated.⁸

Representational content must be internally related to representation in the logical sense of internal relation. A logically *external* relation is one that can change without changing the relata; an internal relation is one that the relata must be in if they are to be what they are. A glass of water on a table is in the external relation of being “on” the table; it could be taken off the table with no change in what it is. An arc of a circle is internally related to the center of that circle: it could not be the arc that it is without having that relation to that center point.

If content is externally related to representation, then, by assumption, there is nothing inherent in the representation that carries or determines that content. The content is arbitrary, whether there even is any content at all is arbitrary, to the purported representation being what it is. The representation is, at best, an arbitrary encoding. This works fine for such representations as Morse code, blueprints, and so on, but the externality of the relation requires that an *interpreter* provide the connection between the representation and its content. That yields the familiar infinite regress of interpreters if it is mental content that we are trying to model. Internally related content is the only way to avoid interpretive homunculi.

The remaining two desiderata are relatively straightforward. Content must be functionally accessible to a system in order to be content for that system at all. Content must be normative for a system in order to be content for that system at all — content is a normative phenomenon.

The dynamic presuppositions of the interactive model are internally related to the interaction indications to which they belong. Those interactions could not be what they are without having those appropriateness conditions. An input processing model, in contrast, generates an internal state that is externally related to the history of the input flow. Assuming that the crucial correspondence relationship is causal, nomological, informational, or evolutionary does not provide an internal relationship between the representation and its content, and, therefore, does not avoid the homunculus regress (e.g., Bickhard & Richie, 1983). Clearly, the interactive model of content is system accessible — the content is internally related to the organization of system organization — and it is system normative — it is normative relative to the processes of recursive self-maintenance of the far from thermodynamic equilibrium system.

Structural isomorphism is internally related to structure, but only if the structure is itself well defined. Wittgenstein’s structural model in the *Tractatus* (Wittgenstein, 1961) involved well defined structure, because it was logically defined structure, but it was neither system accessible nor system normative. Cummins’ structural isomorphism has the advantage of being system accessible, because it is ultimately a physical or biological

“structure”, but it is not system normative, and that makes the identity of the relevant physical or biological structure itself indeterminate, thus not internal, as well as making the distinction between a correct application of the representation to the target and an incorrect application impossible to model for the system itself.

The point was made earlier that content is not accessible for etiological models nor for Fodor’s model: relevant histories and relationships among counterfactual dependencies are not available to the typical organism. There is a claim of normativity in Millikan’s model, though whether there is normativity for the organism is questionable. Certainly there is no causally efficacious normativity for the organism. The appearance of normativity in Fodor’s model derives from the sense in which asymmetric dependency seems to capture something of the parasiticness of error on correctness, but none of this is determinable in the system itself, and, although one side of an asymmetric dependency is picked out just in virtue of the asymmetry, it is unclear why there is any *normative* force in one side of an asymmetric dependency versus the other side, if that is all that is available for making the distinction.⁹

Conclusion

Interactive representation is naturalistically emergent as the solution to the problem of action selection. It is not epiphenomenal, and it emerges naturally in the evolution of biological agents. It has resources with which to model more complex forms of representation. Interactive representation is internally related to its content, and, therefore, does not require an interpreter: it is free of homunculus problems. Interactive representation is inherently related to the functioning of the system — it is certainly functionally accessible. Interactive representation is inherently normative: it has the function of serving action selection, and that function is normative in its contributions to maintaining the organism in its far from equilibrium state. Interactive representation has truth value; it trivially accounts for the possibility of representational error; and it accounts for the possibility of system detectable error, and is thus compatible with the facts of error guided behavior and error guided learning. It also thereby avoids the radical skeptical argument. Interactive representation is a candidate for modeling the fundamental nature of representation.

References

- Aitchison, I. J. R. (1985). Nothing’s Plenty: The vacuum in modern quantum field theory. *Contemporary Physics*, 26(4), 333-391.
- Bickhard, M. H. (1980). *Cognition, Convention, and Communication*. New York: Praeger Publishers.
- Bickhard, M. H. (1988). The Necessity of Possibility and Necessity. Review of Piaget’s *Possibility and Necessity Harvard Educational Review*, 58, No. 4, 502-507.
- Bickhard, M. H. (1993). Representational Content in Humans and Machines. *Journal of Experimental and Theoretical Artificial Intelligence*, 5, 285-333.
- Bickhard, M. H. (1998). Levels of Representationality. *Journal of Experimental and Theoretical Artificial Intelligence*, 10(2), 179-215.
- Bickhard, M. H. (1999). Interaction and Representation. *Theory & Psychology*, 9(4), 435-458.
- Bickhard, M. H. (2000). Emergence. In P. B. Andersen, C. Emmeche, N. O. Finnemann, P. V. Christiansen (Eds.) *Downward Causation*. (322-348). Aarhus, Denmark: University of Aarhus Press.

- Bickhard, M. H. (in press). "Autonomy, Function, and Representation" *Communication and Cognition*.
- Bickhard, M. H., Campbell, R. L. (1989). Interactivism and Genetic Epistemology. *Archives de Psychologie*, 57(221), 99-121.
- Bickhard, M. H., Campbell, R. L. (1992). Some Foundational Questions Concerning Language Studies: With a Focus on Categorical Grammars and Model Theoretic Possible Worlds Semantics. *Journal of Pragmatics*, 17(5/6), 401-433.
- Bickhard, M. H., Richie, D. M. (1983). *On the Nature of Representation: A Case Study of James J. Gibson's Theory of Perception*. New York: Praeger.
- Bickhard, M. H., Terveen, L. (1995). *Foundational Issues in Artificial Intelligence and Cognitive Science — Impasse and Solution*. Amsterdam: Elsevier Scientific.
- Brown, H. R., & Harré, R. (1988). *Philosophical foundations of quantum field theory*. Oxford: Oxford University Press.
- Campbell, D. T. (1990). Levels of Organization, Downward Causation, and the Selection-Theory Approach to Evolutionary Epistemology. In Greenberg, G., & Tobach, E. (Eds.) *Theories of the Evolution of Knowing*. (1-17). Hillsdale, NJ: Erlbaum.
- Campbell, R. L., Bickhard, M. H. (1986). *Knowing Levels and Developmental Stages*. Basel: Karger.
- Chang, C. C., Keisler, H. J. (1990). *Model Theory*. North Holland.
- Christensen, W., Bickhard, M. H. (forthcoming). The Dynamic Emergence of Normative Function.
- Clark, A. (1997). *Being There*. MIT/Bradford.
- Cummins, R. (1996). *Representations, Targets, and Attitudes*. MIT.
- Davies, P. C. W. (1984). Particles Do Not Exist. In S. M. Christensen (Ed.) *Quantum Theory of Gravity*. (66-77). Adam Hilger.
- Dawar, A., Hella, L., Kolaitis, Ph. G. (1995). Implicit Definability and Infinitary Logic in Finite Model Theory. *Proceedings of the 22nd International Colloquium on Automata, Languages, and Programming, ICALP 95, Szeged, Hungary, July 10-11, 1995*. (621-635). New York: Springer-Verlag.
- Dretske, F. I. (1988). *Explaining Behavior*. Cambridge, MA: MIT Press.
- Fodor, J. A. (1990). *A Theory of Content*. Cambridge, MA: MIT Press.
- Fodor, J. A. (1991). Replies. In B. Loewer, G. Rey (Eds.) *Meaning in Mind: Fodor and his critics*. (255-319). Oxford: Blackwell.
- Hella, L., Kolaitis, P. G., Luosto, K. (1994). How to Define a Linear Order on Finite Models. *Proceedings: Symposium on Logic in Computer Science, Paris, France, July 4-7*. Los Alamitos, CA.: IEEE Computer Society Press.
- Hylton, P. (1990). *Russell, Idealism, and the Emergence of Analytic Philosophy*. Oxford.
- Kim, J. (1989). The Myth of Nonreductive Materialism. *Proceedings and Addresses of the American Philosophical Association*, 63, 31-47.
- Kim, J. (1990). Supervenience as a Philosophical Concept. *Metaphilosophy*, 21(1-2), 1-27.
- Kim, J. (1991). Epiphenomenal and Supervenient Causation. In D. M. Rosenthal (Ed.) *The Nature of Mind*. (257-265). Oxford University Press.

- Kim, J. (1992). Multiple Realization and the Metaphysics of Reduction. *Philosophy and Phenomenological Research*, 52, 1-26.
- Kim, J. (1993a). *Supervenience and Mind*. Cambridge University Press.
- Kim, J. (1993b). The Non-Reductivist's Troubles with Mental Causation. In J. Heil, A. Mele (Eds.) *Mental Causation*. (189-210). Oxford University Press.
- Kolaitis, Ph. G. (1990). Implicit Definability on Finite Structures and Unambiguous Computations. In *Proc. 5th IEEE LICS*, 168-180.
- Kolaitis, Ph. G. (manuscript, 1996). Infinitary Logic in Finite Model Theory.
- Levine, A., Bickhard, M. H. (1999). Concepts: Where Fodor Went Wrong. *Philosophical Psychology*, 12(1), 5-23.
- Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT Press.
- Millikan, R. G. (1993). *White Queen Psychology and Other Essays for Alice*. Cambridge, MA: MIT Press.
- Moreno, A., Ruiz-Mirazo, K. (1999). Metabolism and the Problem of its Universalization. *BioSystems*, 49, 45-61.
- Piaget, J. (1987). *Possibility and Necessity. Vols. 1 and 2*. Minneapolis: U. of Minnesota Press.
- Rosenthal, S. B. (1983). Meaning as Habit: Some Systematic Implications of Peirce's Pragmatism. In E. Freeman (Ed.) *The Relevance of Charles Peirce*. (312-327). La Salle, IL: Monist.
- Saunders, S., Brown, H. R. (1991). *The Philosophy of Vacuum*. Oxford University Press.
- Sciama, D. W. (1991). The Physical Significance of the Vacuum State of a Quantum Field. In S. Saunders, H. R. Brown (Eds.) *The Philosophy of Vacuum*. (137-158) Oxford University Press.
- Weinberg, S. (1977). The Search for Unity, Notes for a History of Quantum Field Theory. *Daedalus*, 106(4), 17-35.
- Weinberg, S. (1995). *The Quantum Theory of Fields. Vol. 1. Foundations*. Cambridge.
- Weinberg, S. (1996). *The Quantum Theory of Fields. Vol. II Modern Applications*. Cambridge.
- Wittgenstein, L. (1961). *Tractatus Logico-Philosophicus*. New York: Routledge.

Endnotes

¹ For a more detailed treatment of these issues, see (Bickhard, in press; Christensen & Bickhard, forthcoming).

² For a more detailed treatment of these points, see (Bickhard, 1993; Bickhard & Terveen, 1995).

³ The interactive model is a pragmatic model in the sense of being action based rather than a spectator model (see below), but it is closer to Peirce's model of meaning as anticipatory habit than to his model of representation per se (Rosenthal, 1983). Anticipations can be false, and can be (fallibly) detected to be false.

⁴ I characterize these as "generally" Piagetian because, although one of Piaget's many massive contributions was to construct such action based representations, I don't think the details of his model are all correct (e.g., Bickhard & Campbell, 1989; Campbell & Bickhard, 1986).

⁵ E.g., Bickhard, 1980, 1998; Bickhard & Campbell, 1992; Bickhard & Richie, 1983; Bickhard & Terveen, 1995.

⁶ See Millikan (1984) for this point.

⁷ Note that the claim that such system detectable error checking is impossible is the conclusion of the radical skeptical argument: any such check of a representation is simply re-using the representation again, and, thus, is circular.

⁸ Russell rejected internal relations because of their role in the idealisms of Green and Bradley (Hylton, 1990), and they have largely disappeared from the scene. With respect to representation, it was objected that for a representation to be internally related to what it represented was for mental representation to itself participate in determining the world that it represented: any change in that internal relation would, by assumption, involve a change in the world that was represented. Note, however, that the criterion proposed here is not that of an internal relation between a representation and that which it represents, but, rather, between a representation and its content.

⁹ Needless to say, models that attempt to account for representation in terms of an external observer's gloss on, or explanation of, a system (e.g., Clark, 1997; Dretske, 1988) do not involve system normativity, and generally do not have internally related content. In being explicitly observer dependent, they do not even make a claim of being naturalistic.