Representational Content in Humans and Machines

Mark H. Bickhard

Mark H. Bickhard
Department of Philosophy
Lehigh University
Bethlehem, PA 18015
215-758-3633
MHB0@LEHIGH.EDU

Thanks and acknowledgements are due to many people for comments, suggestions, questions, objections, and discussions that have contributed to this paper. These include: Phil Agre, Bob Barnes, Gordon Bearn, Terry Brown, Donald Campbell, Robert Campbell, Michael Chapman, Ron Chrisley, Bill Clancey, Robert Cooper, Ken Ford, Clark Glymour, Steve Goldman, Richard Kitchener, Ben Kuipers, Ralph Lindgren, Josef Perner, Benny Shanon, Loren Terveen, and, most especially, Norman Melchert. The multitudinous errors and omissions that remain are entirely my responsibility.

Bickhard, M. H. (1993). Representational Content in Humans and Machines. *Journal of Experimental and Theoretical Artificial Intelligence*, *5*, 285-333.

Representational Content in Humans and Machines

Mark H. Bickhard

Abstract

This article focuses on the problem of representational content. Accounting for representational content is the central issue in contemporary naturalism: it is the major remaining task facing a naturalistic conception of the world. Representational content is also the central barrier to contemporary cognitive science and artificial intelligence: it is not possible to understand representation in animals nor to construct machines with genuine representation given current (lack of) understanding of what representation is. An elaborated critique is offered to current approaches to representation, arguing that the basic underlying approach is, at root, logically incoherent, and, thus, that standard approaches are doomed to failure. An alternative model of representation — interactivism — is presented that avoids or solves the problems facing standard approaches. Interactivism is framed by a version of functionalism, and a naturalization of that functionalism completes an outline of a naturalization of representation and representational content.

Key words: artificial intelligence, cognitive science, connectionism, control, emergence, encodingism, epistemology, Fodor, functionalism, idealism, information, innatism, interactivism, narrow content, naturalism, ontology, representation, semantics, skepticism, transduction

Representational Content in Humans and Machines

Mark H. Bickhard

This chapter focuses on the problem of representational content. The problem of representational content is among the most contentious issues in contemporary artificial intelligence, cognitive science, and the philosophy of mind. It is not a new problem, however, and in fact has been around — unsolved — for millennia. What is new are many new conceptual tools and approaches and at least some sense of what went wrong with earlier attempts at solution. Nevertheless, I will argue that contemporary approaches to the problem of representational content are committed — sometimes explicitly, more often implicitly via unexamined presuppositions — to the same fundamental error that has dominated thought about the mind since at least Aristotle. If so, then attempts to understand human or other animal representational content, or to build machines with genuine representational content, are doomed to failure so long as they remain within this flawed framework. To escape from this trap, I offer a fundamentally different alternative conception of the nature of representation — one that does provide a framework for understanding representation in human beings and for building machines with genuine representation. I begin with a little recent historical and philosophical contextualization.

The problem of representational content is a central aspect of the problem of intentionality — of how any system or agent can instantiate any sort of 'aboutness' relationship with its world. Representational content constitutes the system's knowledge of what (a) representation is *supposed* to represent, prior to any questions of warrant or truth value or consciousness of or attitude toward that representation. Representational content is whatever it is that constitutes a representation of a dog as representing a dog rather than as representing something else, or rather than not being representational at all. It is, of course, possible that the problem of representational content cannot be solved independently of solving some of the related problems of intentionality — warrant or truth value or attitude, perhaps.

The problem of representational content is particularly troublesome within the framework of contemporary naturalism — whether functionalism, computationalism, instrumentalism, eliminativism, or any other variety; it is worth keeping in mind, however, that no satisfactory solution has ever been offered even within the framework of a non-naturalism — an ontological dualism, for example. Content yields a particularly uncomfortable problematic for naturalisms, however, since content is central to intentionality, and intentionality, in turn, is the fundamental challenge to naturalism. Mind, especially with respect to its properties of intentionality, seems so fundamentally different from the rest of the natural world that the prima facie case would seem to be for simply accepting that difference in an ontological dualism.

Prior to the relatively modern successes of physics, chemistry, and physiological and evolutionary biology, in fact, some such dualism did seem to be obvious and unavoidable. The naturalistic successes of providing integrated accounts of such phenomena as fire, chemical powers and affinities, life, and so on, however, has switched the contemporary presumption in favor of naturalism. It is now assumed that all phenomena will eventually come under some sort of naturalistic understanding.

Intentionality, and representational content, however, are not only among the earliest and most serious apparent counterexamples to any such hopes for naturalism, they are today also just about the only hold outs that haven't yet succumbed to naturalistic analysis.

Naturalism about representational issues (especially naturalisms of certain computer-compatible forms) has received an enormous boost from the development of computers, and from subsequent progress in artificial intelligence and cognitive science. So many aspects of representation seem to be so powerfully capturable by such systems and within current frameworks that the working presumption is that all aspects will be similarly capturable. There are even naive claims that they have already been captured (Laird, Newell, Rosenbloom, 1987; Newell, 1980a, 1980b; see Bickhard and Terveen, in preparation).

Notoriously, however, that has not yet happened. Recognition that something fundamental remains missing goes by such terms as "the empty symbol problem," (Block, 1980b) or "the symbol grounding problem." (Harnad, 1990) Major battles are fought in the theoretical frontiers of cognitive science and in the philosophy of mind over alternative approaches and proposals in this arena. Progress has been made at least in the sense of reaching relative consensus that certain approaches — like behaviorism — cannot work, but there is no consensus concerning what will work.

If the thesis of this chapter is correct, however, then none of the currently proposed approaches to the problem can possibly work: they all share, with each other and with the last two thousand years of thought in this domain, an underlying logical incoherence in their presuppositions. Specifically, they all propose or presuppose that representation is fundamentally constituted as some form of encodings. Demonstrating that this assumption is necessarily false and incoherent, and providing at least an adumbration of an alternative conception of representation, is the burden of this chapter.

The discussion will proceed in six steps: 1) providing an explication of what encodings are, 2) presenting an initial demonstration of the incoherence of assuming that representation is constituted as encodings, 3) showing that a number of 'alternative' approaches are in fact only variants of this general encodingism framework, 4) developing several corollaries of the basic incoherence argument, both for the purpose of strengthening the force of the conclusion, and to gain further insight into a diagnosis of what is wrong with encodingism, 5) presenting several illustrations of encodingist distortions and false directions in contemporary literature — with J. Fodor as a representative primary source, and 6) presenting an alternative conception of representation and showing that it is not vulnerable to the critiques of encodingism. The discussion then returns to some of the problems vexing contemporary encodingism within the (hopefully) illuminating perspective provided by that alternative model of representation.

What are Encodings?

The nature of encodings will be explicated in terms of the patent paradigm case of Morse code. I will argue that all genuine representational encodings have the same basic character as those of Morse code, and that it is logically incoherent to assume that all representation has this character.

In particular, the property that I want to extract from Morse code is that encodings are stand-ins for what they encode: "**. . .**" stands-in for "**S**", and "**- - -**" stands-in for "**O**". These stand-ins are useful because dots and dashes can be sent over telegraph wires, while "**S**" and "**O**" cannot. Similarly, extremely complex manipulations at fantastic speeds are possible on the bit pattern stand-ins in a computer, but not on characters. In general, encoding stand-ins change the form and the medium of representations because of differing potentialities of manipulation in the differing forms and media.

The critical point, however, is that such stand-ins represent whatever they represent — carry whatever representational content that they carry — by virtue of having borrowed it from whatever they are standing-in for. The stand-in relationship **is** a relationship of representational content transfer. But, in order for such a transfer to occur, in order for an encoding to be defined, some already existing representation (or string of representations) must already carry the desired representational content. Encodings are defined in terms of already present representations, and cannot **be** representations, cannot carry any representational content, except via such stand-in relationships.

This is, of course, not problematic for genuine encodings, such as Morse code or computer codes. The stand-in relationships are explicitly defined by the designers or users. In fact, such definitions can iterate for any finite number of levels: "**X**" can be defined in terms of "**Y**", while "**Y**" is defined in terms of "**Z**", for example. They cannot iterate unboundedly, however: that would require an unbounded regress of actual encoding relationships in order for the top levels to carry any representational content at all.

The Incoherence of Encodingism

The consequence is that any such definitional levels must be finite in number, and, therefore, that there must be a lowest level. It is at this presumed lowest level that we encounter impossibilities of principle in the assumption that this lowest level is itself constituted as encodings. In particular, if it is assumed that representation is intrinsically constituted as encodings, then this lowest level *must* be itself constituted as encodings. But, if that encodingist assumption regarding the lowest level of representations is impossible, then encodingism itself is impossible. In presupposing an impossibility — that the lowest level of representations must themselves be encodings — encodingism renders itself logically incoherent. This is what I wish to demonstrate.

Any such element of the lowest level of encodings must, by assumption, not be defined in terms of any other representations, else it would not be at the lowest level. It must be logically independent of any other representations. Yet it must carry some representational content in order to be an encoding at all. The impossibility arises when this issue of representational content at the lowest level of encodings is examined.

Consider any element of this presumed lowest level, say "**X**". "**X**" cannot be defined in terms of any other representations — it cannot obtain its representational content from any other representations — by assumption; but "**X**" must nevertheless carry some representational content. The only possible source of that content is "**X**" itself, which leaves us with: " "**X**" represents whatever it is that "**X**" represents" or " "**X**" stands-in for "**X**" " But this does not suffice to provide "**X**" with any representational content at all. It leaves "**X**" representationally empty, therefore not an encoding at all, *contrary to assumption*. A strict encodingism presupposes that "**X**" carries representational content, yet makes it impossible for "**X**" to carry any content. The presuppositions of encodingism have forced a logical contradiction — encodingism is logically incoherent.

Variants of Encodingism

The stand-in perspective on encodings derives very readily from actual encodings such as Morse code. When encodings are invoked to do epistemic work, however, such as in encoding conceptions of perception, cognition, or language, it is not usually Morse code that is appealed to. In this section, I will examine several alternative conceptions of encodings and argue that they are in fact all variants of each other — in particular, that they are all variants of the stand-in version. One consequence of this is that encodings are incapable of accomplishing *any* of the epistemic tasks for which they are commonly invoked.

User and Designer Semantics. Among the most common versions of encodings are those constructed by *users or designers* of computer and related systems. In such instances, some element, or string of elements, is designed or stipulated to represent something else, say **Y**, where "**Y**" will be some other (string of) elements, usually in English or some other natural language. In effect, the encoding is defined by specifying what it is to be taken to represent. The form of such user or designer stipulations is slightly different from that of the stand-in definition, but the user/designer form is just a use-mention variant of the stand-in form: The user or designer definition in terms of what the encoding is to be taken to represent has the form

"**X**" represents **Y**

while the stand-in definition has the form

"**X**" stands-in for "**Y**".

The shift from "represents" to "stands-in for" as the defining relation is merely the adjoint for the shift from use of "**Y**" to mention of "**Y**". In both cases, "**X**" is being defined in terms of "**Y**", and, in both cases, whatever representational content "**X**" carries is provided by "**Y**". An encoding defined by what it is to represent, then, is just a minor perspective change on an encoding defined as a stand-in. In particular, a representation defined by what it represents **is** an encoding.

A user or designer semantics for symbols or symbol strings in a computational system is a semantics — an encoding semantics — *only* for the user or designer. The defining representational/stand-in relationships are known only by the user or designer, not by the system. Even the *existence* of any such relationships is not known by the system.

Observer Semantics. In the case of computational systems, if we conclude that the 'symbols' in the system have no representational content, are not representations, for the system itself, we will also generally conclude that *the system has no representations*, no knowledge, at all: if not the 'symbols', then there is little plausible alternative. In general, however, the question of whether or not a system itself knows anything at all is an *additional* question to those of whether or not the system under consideration has any knowledge concerning such encoding-defining or stand-in relationships, or even concerning the existence of any such relationships.

The question of whether the system itself knows anything at all — of whether the system itself is any kind of epistemic system at all — becomes most important when considering living systems: whether or not it makes sense to claim that a computer system knows anything at all turns primarily on whether or not it makes sense to claim that the symbol strings in the system represent anything at all *for the system itself*. Since neither the existence nor the content of any defining relations for those symbols can be known by such a system, the two questions fail together. For living systems, however, they may not fail together. In particular, we might examine the sensory system of frogs, for example, and conclude that certain patterns of retinal or brain activity represent corresponding spatial and temporal patterns of light, on the basis of which the frog flicks its tongue in order to eat a fly. In a case like this, we do not generally raise the contextual question of whether or not frogs have any representations at all — we assume that they do — though we may be quite interested in what *sorts* of representations the frog does have.

Nevertheless, in claiming that particular neural activity patterns encode certain light patterns, we are committing the error of assigning an *observer semantics* to the system being observed. It is the *observer* who notices the correspondence between light patterns and neural activity patterns, and it is the observer who uses that correspondence to define an encoding representational relationship. But the frog knows nothing of any such relationships, nor their content — nothing about light patterns at all, in fact. The relevant representational content is not present, and could not even possibly be present.

Such an observer semantics is a second variant of encodingism: it is just a version of a user or designer semantics. In both cases, an epistemic agent outside of the system under consideration defines an encoding relationship that exists only for that outside observer. In the user and designer cases, that definition is relatively arbitrary, while in the case of an observer of an organism that is already granted status as some sort of epistemic agent itself, that definition may be based on observed correspondences and covariations between activities internal to the system and events or conditions outside of the system. Because of the foundation of such encoding definitions on correspondences that are actually observed between the system and its environment, there is not the arbitrariness that is to be found in user or designer semantics, and there is, correspondingly, a temptation to conclude that these definitions are somehow inherent in the system itself, and, therefore, that they constitute encoding representations for the system itself.[1]

Nevertheless, it remains the case that the very existence of any such correspondences, and the "other ends" of any such correspondences (e.g., light patterns) are in no way represented in the system (e.g., the frog) under consideration. Presumably the activities of the frog must in some ways be sensitive and responsive to the events and conditions in its environment, and presumably such causal relationships and consequent factual correspondences between retinal patterns and light patterns, and between light patterns and flies, will play important roles in insuring appropriate sensitivity and responsiveness. But the question of representational content for the systems themselves — for the frog — has not even been addressed in such analyses.

Such analyses, at best, provide part of a *functional* analysis of how the system — frog — manages to survive in evolutionary ecological conditions. Such functional analyses can be interesting and important, but they are *not the same thing as epistemic or representational analyses*. They are aspects of functional analyses of the system's interactions with its environments, of the *informational* functional relationships, where 'informational' is understood in the sense of manifesting correspondences and covariations. Such analyses may, or may not, contribute to epistemic analyses, but they are not themselves epistemic analyses.

In general, the most seductive paths to encodingism derive from various versions of this assumption that correspondences and covariations between systems and their environments — as noted by an observer — can constitute encodings, not just for the observer, but for the systems themselves. Such factual correspondences, whether directly causal or more generally informational, are taken to be the core of encodings in most computational approaches today. Several problems with such approaches are recognized in contemporary literature, but they are taken to be subsidiary problems of detail, to be corrected from within the basic correspondence-as-encoding framework, not as symptoms of deeper problems with the whole approach. I will examine some of these internal difficulties and purported solutions within the correspondence-as-encoding approaches later.

Transducer Semantics. First, however, I would like to illustrate the range and variety of perspectives on cognition and representation that commit this correspondence-as-encoding error, and, thus, are committed to the logical incoherence of encodingism in general. The illustrative example used above was that of a frog. But exactly the same error is commonly committed in attempts to model human and other sensory systems. Quite often, sensory processes are straightforwardly examined in terms of "sensory encodings", where sensory encodings are analyzed in terms of factual correspondences between neural activity and environmental events and conditions (e.g., Carlson, 1986).

A common terminology is that of "transduction": sensory receptors are taken to "transduce" environmental stimuli into internal encodings for those stimulus conditions, on the basis of which the brain is assumed to infer or construct encodings of the world in general (Fodor and Pylyshyn, 1981). Transduction is, strictly, the transformation of one form of energy into some other form of energy. In examining the processes of retinal light reception, then, the notion of transduction is quite appropriate. Furthermore, such transduction relationships can certainly provide instances of the factual correspondences that encodingists look for. But to conclude that any such factual/lawful transduction correspondence relationship constitutes an epistemic relationship is simply a non-sequitur. Correspondences are cheap — they are everywhere — whether factual, lawful, informational or any other kind. Every lawful regularity in the universe provides unbounded classes of instances.[2] Representation must at least be more than that, if not something different altogether. Factual correspondence relationships are not ipso facto epistemic relationships.

The seductive power of correspondence-as-encoding is impressive. I illustrate with a straightforward logical non-sequitur that is the result of this seduction. In purportedly demolishing Gibson's theory of perception (Gibson, 1966, 1977, 1979), Fodor and Pylyshyn (1981) not only conclude that retinal transduction is an encoding process, they also conclude that it *must* be. Both conclusions are non-sequiturs, but the latter conclusion is particularly illustrative of the manner in which encodingism presuppositions circularly support themselves. Fodor and Pylyshyn use information in the covariation sense, and point out that the visual system must pick up information — must 'pick up' the factual status of being correlated — from the environment in order to be functionally useful at all. So far, so good. But they then slide into the position that the visual system must, therefore, transduce "that the light is so-and-so." (Fodor and Pylyshyn, 1981, p. 165). This is a shift from the point that a system, in order to respond appropriately to its environment, must somehow differentiate internal states that are in some sort of factual correspondences with those environments, into a claim that *the only way to do that* is to encode — "transduce" — that the environment is "so-and-so". But to encode — "transduce" — that the environment is so-and-so is to have representational content that the environment is so-and-so ("implicitly" so the story goes). This is an egregious and illustrative non-sequitur.

In the first place, constructing internal conditions that are in *factual* correspondence with environmental conditions can certainly be functionally useful for the organism, and such correspondences are in fact given by the correct energy-transformation version of the notion of transduction; but *nothing* in this story gives or requires representational content. Thus, the non-sequitur is egregious. Furthermore, the underlying presupposition of the non-sequitur is that the only way to pick up information about the environment — to construct internal states in factual correspondence with external states — is to encode those external states. It is the encodingist presupposition itself, then, that underlies the non-sequitur encodingist conclusion. In fact, *any* internal process yielding environmentally covariational internal states will construct covariational informational internal states — information, in this sense, simply *is* covariation, not representation. Variants of this circularity are inherent in encodingism. The incoherence argument, and the corollaries to be examined later, are all versions of such circularities. Thus, the non-sequitur is illustrative. (A more detailed analysis of this and other errors in Fodor and Pylyshyn, 1981 — as well as a non-encodingist framework for perception — can be found in Bickhard and Richie, 1983.)

Notions of cognition and representation are dominated by user, designer, and observer semantics. Transduction is an example of an attempt to add something — causality, perhaps nomological-ness — to simple correspondence in order to make it into encoding. It doesn't work; strictly, it doesn't even address the issue of representational content for the system itself. The analyses are done strictly from the observer perspective.

Connectionism and PDP. Another example is provided by Parallel Distributed Processing or Connectionist approaches (Horgan and Tienson, 1988; McClelland and Rumelhart, 1986; Rumelhart and McClelland, 1986; Waltz and Feldman, 1988). Simply put, a PDP network that has "learned" to categorize input patterns is construed as constituting a representer of those categories: specifically, the resultant pattern of node activations given some input pattern is construed as an encoding representation of that input pattern. The factual correspondence between the input categories and the node activation patterns is not stipulated, as in user semantics, and it is not tempting to construe it as causally nomological, as with sensory transductions — instead, it is "trained" into the weights of the network. Nevertheless, it is still no more than a factual correspondence with no epistemic content for the system itself. The ability to "train" such correspondence categorizations may be of fundamental importance for some purposes, but it does not address the fundamental issues of representational content any more than does "transduction". Adding a "training" origin to correspondence works no better than adding causality or lawfulness.

Analogical Semantics. Still another variant is to assume that, although there are serious problems with *symbolic* encodings, nevertheless, *analogical* correspondences do constitute representations (Harnad, 1990). But the shift from digital to analogical correspondences has no bearing whatsoever on the basic issue of whether or not the system itself has any knowledge of, any representational content for, either the existence or the content of those correspondences. If the system doesn't know that there is any such correspondence, then, no matter how factual, lawful, informational, trained, or analogical it might be, the correspondence cannot constitute an encoded representation for that system. Similarly, if the system doesn't know what the correspondence is with, it cannot constitute an encoded representation for that system. On the other hand, if the system does know what the correspondence is with, then the system does have an encoding — but the system has to already know *that which* the correspondence is with in order to know that *that* is what the correspondence is with. The system must already have the representational content for 'dog' in order for any correspondence to exist between "dog" or "canine" and dog. That is, the system must *already have* the relevant representational content in order to have the *encoding* for that content. The encodingism circularity shows up here as: you must already have representation in order to get representation.

Encodings and Epistemic Boundaries. Encodings, then, can never be the ground for new representational content. They are incapable of representing anything for which the encoding relationship itself, including the representational content, is not already known. As mentioned before, encodings can, nevertheless, be extremely useful. But encodings cannot be useful for providing *new* representation, only for changing the form of representation already available. In particular, encodings cannot cross the boundaries of epistemic agents to provide grounding representation of anything outside of those agents — perception: any such representational knowledge must already exist inside the agent in order for the encoding to be defined. Similarly, encodings cannot cross from outside of an agent to properties internal to that agent — language — for the same reason: representations would already have to exist of those internal properties in order for the external encodings to be defined or understood. So, encodings cannot cross the boundaries of epistemic agents in either direction, because to do so is to provide new representation of whatever is on the other side of that boundary, and that is precisely what encodings cannot do.

The consequence of this, however, is that encodings cannot cross from mind to environment in perception, or from environment into mind in the form of decoding expressions (encoded utterances) of mental contents. The core of the problems of epistemology is that of new knowledge, new representation, but new representation is precisely what encodings cannot account for: encodings cannot perform ***any*** of the basic epistemological tasks for which they are standardly invoked. Among other consequences, the presumed information processing sequence — encoded representation processing sequence — of perception, cognition, and language, cannot be correct (Bickhard, 1992a).

Corollaries of Incoherence

The incoherence argument is only one of a whole class of corollaries, some ancient, some new, regarding the encodingist notion of representation. Perhaps the oldest argument in this class is that of skepticism.

Skepticism. The skeptical argument turns on the point that, in order to check to see if an encoding representation is correct, it must be checked against that which it is taken to represent, but, under the encodingist assumption, that encoding itself is the only epistemic access available to what it is taken to represent. Therefore, any such check is circular. A slightly more sophisticated version is to construe the checking of an encoding in terms of its supposed consequences, but those consequences too are epistemically accessible only via questionable encodings. So the "checking" is again circular, this time at the level of the encoding system rather than a single encoding. Thus, all such representational checking — so long as representations are construed as encodings — is circular, and there is no ground or warrant for our representations (Burnyeat, 1983; Popkin, 1979).

Idealism. One conclusion that might be drawn from this impossibility argument — so long as it is taken as applying to all representation — is to conclude that it is superfluous to postulate anything on the other end of the encoding relationship at all. We can never epistemically access anything on that other end, all we ever really have are the encodings themselves. So, the conclusion goes, it is a simple application of Occam's razor to conclude that our world is in fact nothing more than our representations. That is, a recognition of the insolubility of skepticism can readily yield an idealism (e.g., Hegel; Pippin, 1989).

But such a move to idealism makes sense only insofar as the skeptical argument is itself accepted universally. I will argue that the skeptical argument is valid only for encoding representations, and not for the alternative model of representation that I will present. If so, then idealism, like the skepticism that gives rise to it, is equally committed to an underlying presupposition of encodingism.

The Copy Argument. Still another variant on the encodingism circularity concerns the issue of origin rather than accuracy. Basically, the system must already know what its world is comprised of in order to construct encodings of that world — encodings must already have representation in order to get representation. One way of putting this is to point out that, if our representations of our world are copies of the world, then we must already know the world in order to construct our copies of it (Piaget, 1970).

Skepticism and idealism result from asking how we can check our encodings; the copy argument from asking how we can know what encodings to generate or use. The incoherence circularity arises from asking how a system could know what its representations are even *supposed* to represent, prior to any issues of construction or accuracy. Within encodingism, there is no answer.

Substance Ontologies and Emergence. Perhaps the deepest perspective on these issues is an *ontological* perspective. Virtually all sciences have had an early historical phase in which the basic subject matter of the science was presumed to be some sort of substance. Such substance ontologies have almost universally been abandoned: we no longer accept phlogiston theories of fire, or caloric theories of heat, or magnetic fluid theories of magnetism, or vital fluid theories of life, and so on. In all cases, these have been superseded by process theories. The move from substance ontologies to process ontologies has been almost universal among sciences.

Substance ontologies come in several varieties. They all are grounded in the postulation of some set of basic substances. In one version, these basic substances are assumed to be infinitely divisible, in which case the basic form of constitution of the world will be in terms of blends of the basic substances. The ancient Greek's earth, air, fire, and water, for example, had this character. In a second version, the basic substances are understood to be composed of indivisible particles — atoms — in which case the form of constitution will be combinations of those atoms. This version has ancient precedents as well.

Such forms of constitution, in turn, can be understood in terms of unstructured aggregations, or in terms of rigid structures, in which the components somehow lock into place. Blends of divisible substances are not usually construed as involving structure, while combinations of atoms are not usually construed as being simply aggregates. Logically, however, all are possible.

Examples of models attempting to explain phenomena in terms of blends of primary types and of models involving presupposed structuralisms are not difficult to find in contemporary literature (see Bickhard and Christopher, under review). My focus here, however, is on atomistic substance approaches, since they are the dominant approach to representation. In particular, standard encodingism *is* an atomism of representation.

There are several critical questions that cannot be addressed from within a substance ontology. These include questions concerning the nature of the differences among the basic substances, questions concerning the stability of the substances, questions concerning the indivisibility or structural rigidity of the atoms and structures, and, most important for my current purposes, questions concerning the origin of the substances. All of these questions, if they are to be addressed at all, require escaping or transcending the basic substance framework. They are questions *about* that framework, not problems that can be solved within that framework.

All explanations within a substance framework appeal to joint contributions of basic substances, whether via combinations of atoms or blends of "stuff". For an atomistic substance approach, explanation is restricted to a combinatorialism of whatever the basic atoms are taken to be.

This is as true for a representational atomism — an encodingism — as for any other atomism. In particular, all representation within an encodingist model is to be "explained", if at all, in terms of combinations of atomic encodings: encodingism forces combinatorialism.

But the impossibility of accounting for the basic substances in terms of combinations of those basic substances is just the general ontological version of the incoherence problem — the *atomic* (substance) encodings themselves cannot be accounted for in terms of combinations of those basic (substance) atomic encodings. Atoms are not constituted out of atoms. A substance ontology cannot account for its own ontology.

With respect to representation, this translates into: encodingism cannot account for the origin of its own basic encodings. It can only provide combinations of atomic encodings that are presumed to already exist. It cannot model the emergence of representation out of phenomena that are themselves not already representational.

Representation, presumably, *has* emerged at least once since the Big Bang — therefore, encodingism *cannot* be the whole story. Encodingism makes any such emergence impossible. Human beings and other animals, therefore, insofar as they count as proof that representation has emerged, also count as counterexamples to encodingism. If encodingism were correct, then representation would be impossible.

A substance approach to representation presupposes the existence of its basic substances, and cannot account for them. Encodingism presupposes that representation is fundamentally constituted as (combinations of) atomic representations, and thereby presupposes the existence of representation in claiming to account for representation. It is this circularity that plays itself out in all the various corollaries of the incoherence argument, several of which have been presented here.

Innatism. This point has been dimly seen in Fodor's argument that, since we have no way for new basic concepts to be learned or developed, all basic concepts must be innate (Fodor, 1975, 1981). Fodor, however, puts the burden of the problem on theories of learning rather than on theories of representation, and he fails to recognize that the basic problem is logical, not simply theoretical, and therefore that evolution cannot solve it any more than can learning or development. If atomic encodings cannot come into being, if representation cannot emerge out of non-representation, then it cannot do so in evolution either. In this respect, Fodor's radical innatism is another non-sequitur (Bickhard, 1991d; Campbell and Bickhard, 1987).

Connectionism Again. This general point is somewhat more subtle in the case of PDP or connectionist systems. In such systems, the correspondences that are taken to be representations are not apriori designed or defined or "transduced", but, instead, are trained or learned. They seem to be emergent, in contradiction to the general point concerning substance ontologies being unable to address the origins of the substances.

That this is only superficially apparent, however, becomes clear once it is realized that what is trained or learned — what is "emergent" — is not representation at all, but, rather, correspondence. A network 'learns' correspondences with various classes of input patterns, but, as before, it does not learn *that* they are correspondences, nor what those correspondences are with; such correspondences per se do not constitute representations.

To take them as representations is to inject an unacknowledged observer's knowledge of the existence of such correspondences and of what those correspondences are correspondences with. As usual, encodingism requires that representation be already present in order to get representation. In this case, the representation that is already present is observer representation; the observer is the source of the representational content — of the representational substances or atoms. PDP systems have not focused on combinatorial issues regarding those atoms — and there is controversy concerning whether and in what form they can (Fodor and Pylyshyn, 1988) — but the basic encodingist critique, the basic inability of a substance ontology to account for its own basic substances, already applies at the level of taking the network correspondences as constituting representations.

Encodingist Distortions, Blind Alleys, and Red Herrings

Encodingism has not so much been taken as having solved the problems of representation as it has been taken as the only possibility for representation. Difficulties that various encodingist approaches have encountered, therefore, have generally been understood as undermining that particular sub-approach, not as undermining encodingism as a general program. Quite often, however, these supposed subsidiary problems, presumed to be solvable within an encodingism, are themselves *products* of encodingism, and are consequently distorting and misleading as guides to modeling and understanding representation.

Too Many Correspondences. One problem that has been recognized for the correspondence as encoding approach, for example, is that whenever factual correspondences exist — say between retinal activity and various surfaces and edges in the environment — there will also exist an unbounded number of alternative correspondences — with light patterns, with quantum events in the retinal cells, with processes among the electron orbitals in the surfaces of the materials in the environment, and so on. If correspondence is encoding, then which of these correspondences is the relevant one, which is the *encoding* correspondence?

One approach to a solution to this problem is to assume that instances of such correspondences are followed by various activities of the system, and that those activities will be functionally appropriate to only one of the chain or lattice of correspondences. In this manner, correspondence-plus-functionality serves to select from within that class of correspondences the one that is in fact being represented (Bogdan, 1988a, 1988b, 1989; Smith, 1985, 1987). Unfortunately, this entire analysis is still being carried out from within an observer perspective. Such functionality might inform such an observer about which correspondence is the functionally relevant one for the system, but this does not provide the system itself with any knowledge of the existence of any such correspondence nor of the other end of the correspondence. The supposed selection by the functionality among the class of correspondences in order to specify which of them is to be represented does not render that correspondence, or any other, an epistemic relationship. Correspondence plus functionality does not work any better than correspondence plus lawfulness or trainedness. Such a selection would suffice *only* if all of the correspondences *were already representations*, and all that had to be selected according to function were which of those already available representations was to be the relevant one. Again, encodingism requires that we already have representation in order to get representation.

Another approach to this "multiplicity of correspondences" problem is to look to the evolutionary selection pressures that have selected for the functionality based on the correspondences as means of sorting out which of the correspondences is the relevant one. By now it should be clear that, however much this succeeds, or doesn't succeed, in selecting the relevant correspondence, it will at best select the *functionally* relevant correspondence — the correspondence that can explain why the system activity is functional for the system — from within an observer perspective; but, as is by now familiar, it does nothing toward explicating encoding representational content for the system itself.

Representational Error: The Very Possibility. Still another problem that has much exercised correspondence encodingists is the problem of error: if correspondence is encoding, then how can any correspondence, so long as it exists at all, be in error? For example, if "**X**" is evoked in correspondence to cows, and, presumably, represents cows, but is also on occasion evoked by horses — on dark nights, say — in what sense is the horse evocation in error? Why doesn't "**X**" simply represent "cows or horses" since that seems to be the correspondence class — in which case the horse evocations would *not* be in error.

Fodor proposes an asymmetric dependence criterion for distinguishing the correct evocations from the errorful evocations. Roughly, the idea is that whenever horses evoke cow encodings, they do so parasitically on the (possibility of) cow evocations, and that parasiticness is what distinguishes the correct evocations from the errors. In other words, if cows did not evoke "**X**", then horses wouldn't either, while if horses didn't evoke "**X**", cows would still do so. Horse evocations are dependent on cow evocations, but not the other way around — the dependency is asymmetric (Fodor, 1987a, 1990).

Presumably, something about this intuition must be correct — there must be an asymmetry between correct representation and error. Whether Fodor's criterion works, however, is still questionable: a control molecule fitting into its receptor on a cell surface will trigger internal cell processes that are in correspondence with the molecule, while a poison molecule partially mimicking the control molecule will also fit into the receptor on the cell surface, and will do so in an asymmetrically dependent way. Yet there is no representation, no epistemic encoding, involved at all. Fodor's criterion may (or may not) differentiate correct evocations from error, but it does so on a strictly functional level of analysis, not on an epistemic level. When applied to correspondence encodings, should such exist, it may correctly differentiate error instances, but the encodings must already be there *as encodings* in order for the criterion to pick out encoding errors rather than simple functional parasitisms.

More deeply, even if restricted to encodings, so that the criterion does pick out errors in encoding evocations, the entire analysis is strictly from within an observer perspective on the system, not from within the system perspective itself, and, thus does not address *any* issues of representation for the system itself. The errorful-correspondence encoding problem is a problem within the *observer semantics* for the system, not a problem, nor the solution to a problem, for the system. In the alternative approach to representation that I will present, in fact, the possibility of error is trivial — there is no problematic at all in accounting for the possibility of errorful representation. If so, then the problem of error is simply a red herring produced by the internal complexities and impossibilities of an incoherent encodingism.

It's All Just Observer Semantics Anyway. The problem of multiple correspondences and the problem of errorful correspondences, then, arise only because of the prior acceptance of the correspondence-as-encoding approach. That approach is already restricted to an observer semantics, so any supposed solutions to these subproblems within that approach will at best be solutions within an observer semantics — and not only *within* an observer semantics, but for problems *for* an observer semantics: any such solutions will be solutions to problems that exist only for an observer perspective, not problems for issues of representation per se. They do not provide any account of representation, of representational content, for the system. These problems, then, are distortions and blind alleys with respect to the general problem of representation. Solving them, even if accomplished, even if possible, will not help to solve the primary problem of representation.

Information is not Representation. Oddly, Fodor seems at certain points to partially recognize something like this. In a discussion of Barwise and Perry's situation semantics, he points out that not all situations that contain information (in the correspondence or covariation sense) encode that information — "not every situation encodes the information that it contains". Furthermore, he acknowledges that "we haven't got a ghost of a Naturalistic theory about" encodings (Fodor, 1987b, p. 87), and "... of the semanticity of mental representations we have, as things now stand, no adequate account." (Fodor, 1990, p. 28).[3] This is part of a set of distinctions that I would very much like to impress upon Fodor and all other encodingists: not only is correspondence or covariation information not the same as representation, but representation is not the same as encoding. Furthermore, encodings are a derivative form of representation from a more basic form; presuppositions to the contrary — encoding*isms* — are logically incoherent.

Interactivism: A model of emergent representation.

In the course of the discussion so far, I have issued a number of promissory notes concerning a non-encoding model of representation; in this section, I will limn how to make good on those notes.

Functional Analysis. The model that I propose is developed within control theory, and control theory, in turn, is a framework for a particular kind of functional analysis. Broadly, then, what I wish to propose is a functional analysis of emergent representation.

Contemporary Functionalism. In a general sense, this is convergent with functionalism within contemporary studies of the mind (Block, 1980a; Fodor, 1975; Newell, 1980a). More careful consideration, however, reveals fundamental differences. Contemporary functional analysis might, from the perspective I will using, be better called "Formal functional analysis plus representation." The point of the qualifier "*formal* functional analysis" is that contemporary functionalism is a functionalism of formal process, and formal process only. Its foundational mathematics is that of automata theory and Turing machine theory.

There are two senses in which the notion of "formal" applies here: one is that formal processes are characterized only up to the *sequence* of the steps in the process. There is no issue of time or timing beyond formal sequence in such a framework: the formal properties of a Turing machine depend not at all on the timing of the steps in its calculations. In this respect, the model I propose requires a notion of process, and of functional and control organizations of process, that both transcends and is more powerful than Turing machine theory: it requires a model of real time, and real timing, processes (Bickhard and Terveen, in preparation). The basic argument is that the interactive model requires real time successful interactions with an environment, and formal functionalism cannot model such considerations in a non—ad hoc manner. In this discussion, however, I will not address issues of timing, and will proceed instead within a framework that is explicable within automata theory.

The second sense of "formal" that is intended in the characterization of contemporary functionalism is that formal processes are construed as operating on formal encoded symbols. Exactly what formal symbols are, or what operating on formal symbols is, is a matter of dispute and exploration. Roughly, for Fodor, the notion is that causal processes operate on physical instances of symbols, whatever those instances may be, only in terms of the causally relevant properties of those instances of symbols — in particular, only on their "shape" or their non-semantic properties in some sense. Underlying this notion is a basic commitment to naturalism: whatever formal symbols are, and whatever formal processes may be, they should be naturalistically explicable.

This sense of "formal" connects to the "plus representation" in the characterization of contemporary functionalism, in that some sort of model of representation is *presupposed* in this framework — the symbols — while the emergent origin of representation is precisely what I take as the *central* problem to be addressed. With regard to the problem of representation, this presupposition constitutes a circularity. More particularly, some sort of *encodingism* is presupposed in what is standardly called functionalism, and it is by now clear that I regard that as unacceptable — incoherent, to be precise. So, I need a functionalism that does not involve any presuppositions regarding representation (and that is competent to model timing considerations; see Bickhard and Terveen, in preparation).

Consequence. The general notion of function that I wish to rely on — a broader notion than that of control per se — explicates function in terms of consequence: If **A** exerts influence on, has consequences for, **B**, then the functions of **A** relative to **B** are the effects that **A** has for **B**. It may be the case that substitutes for **A** could serve the same function for **B**, in the sense of having the same consequences.[4] If **A** is a complex control structure, in particular, there may be many alternative such structures that could serve the same control function for **B**. The importance of such notions of function, and the sense in which the framework that I am developing is a functional framework, is that physical level system process models can be replaced by models of the *consequences*, the functions, of those processes for other processes. This is the sense in which a functional perspective abstracts away from particularities of realization.

Control In particular, the explication of a control relationship is itself an explication in terms of consequence — a consequence of process influence, or control — and is, therefore, a functional explication. Consider two (sub)systems, **A** and **B**, engaging in physically specifiable processes. If the course or the outcome of the processes in **A** influence the course of the processes in **B**, then I wish to say that **A** exerts *control* on **B**. A stable relationship of such control constitutes a *control relationship*. Such control relationships involve differences in **A** processes evoking or triggering or selecting or *differentiating* differences in **B** processes. Such power of selection or differentiation of **B** processes by **A** can be no greater than the potential variety in **A** processes — it is only that variety of differences in **A** that is available to induce differences in **B** (Ashby, 1960). Classic mathematical information theory provides a measure of such variety, and, thus, of such potential control of one system on another.

Control Structure. A stable organization of control relationships among various (sub)systems constitutes a *control structure*. Note that a limiting case of a control relationship is a switching relationship, so the notion of a control structure offers at least the power of switching theory, and anything that can be constructed on the basis of switching theory — e.g., a Universal Turing Machine.

Control Flow. If the control relationships in a control structure include those of switching various components on and off — if there are conditions of process quiescence into which and out of which components can be switched — then the flow of conditions of being switched-on through the system control structure constitutes a *control flow*. Note that an inactive component cannot, by assumption, exert control, so only active — switched on — components can switch on other components. Control flow, then, is a flow of activation of activity, of process, in the system structure. Note that not all control influences constitute control flows: one subsystem can exert control on the processes of another without necessarily switching it on or off. The extreme version of a control flow organization is in a von Neumann computer architecture, in which only one instruction can be active at a time: the flow of activity, of exerting control on the computer processing, from instruction to instruction is a control flow.

Functional Indication. If some condition in a system, say **Q**, exerts control on the switching processes in another, say **S**, such that **S** will be *able* to switch to some third component, say **T**, when condition **Q** holds, then **Q** is an *indicator* of **T** for **S**. For **Q** to indicate **T** for **S**, then, is for **Q** to enable the *possibility* of **S** switching to **T**. Whether **S** actually switches to **T** may depend on many other aspects of process. **Q**, then, is *sufficient* for the *possibility* of **S** switching to **T**, but does not necessarily yield the actual flow to **T**. An indicator enables a switch to switch to particular outcomes. Still another perspective on indication is to note that an indicator for **T** switches the processes in **S** such that switching to **T** is now a possible outcome of the processes in **S**. Note that a control relationship typically can be realized as a causal relationship, but an indicator relationship cannot: an indicator is neither necessary to what it indicates (there could be other indicators of the same possibility) nor sufficient to what it indicates (**S** could nevertheless end up in some other outcome, even though **T** has been indicatively enabled). An indicator is sufficient to the *possibility* of (switching to) what it indicates.

Open Systems. A system that is *necessarily* in interaction with an environment — a physical necessity, at least — is an *open system* (Nicolis and Prigogine, 1977, 1989; Prigogine, 1980). This occurs if the continued existence of the system is dependent on such interaction, such as for a flame: to close a flame off from its environment is to extinguish it. In the case of a control system, an open control system is one that receives control influences from an environment and exerts control on that environment. It is within the conceptual framework of such open system interactive control structures that the explication of representation will proceed.

Goal Directedness. One more notion is needed first, however, and that is *goal directedness*. This is potentially problematic because the prima facie obvious way in which to understand a goal is in terms of a *representation* of the goal conditions to be met. Reliance on such a representational notion of goal would render any explication of representation in terms of it circular, so I need a more primitive, non-representational, conception of goal. What I will need is an organization of control in which goal satisfaction, or lack thereof, can be modeled strictly on internal functional conditions, not on representations of external conditions. If **A** is a switch that either switches control flow to **B** — triggers execution of **B**, which then may return to **A** — or switches control flow elsewhere out of the **A-B** subsystem, and if the internal conditions of **A** that determine that switching are controlled by — influenced by — the environment, then the

A-B subsystem will constitute a goal directed subsystem in the required sense. **A** is simply a conditional, environmentally conditional, switch — either to **B**, or out of the **A-B** subsystem.

This, of course, is even clearer if **B** has a propensity to produce conditions that induce **A** to switch out of the system, but I am going to be more concerned with the sense in which the switching of **A** constitutes functional information (a source of control influence, not representation) in the overall system concerning *whether or not* its switching conditions have in fact been satisfied, than with **B**'s actual propensities to satisfy them. Roughly, if **B** fails to do so (to satisfy the switching conditions), then something (see below) is falsified, and the switching of **A** informationally (covariationally; functionally) captures that. A relationship of control from the environment to **A** is needed here, but that relationship can be modeled in a purely functional manner, and no notion of representation is required. (This is similar to the classic TOTE model, but with a strictly control notion of goal, and without any goal representation; Miller, Galanter, Pribram, 1960.) Once representation is available, however, there is no restriction preventing those switching relationships in **A** from depending on satisfactions of representations — but, again, such representational relationships are not required.

Such a notion of goal directedness, then, is abstracted from the standard representational version of goal directedness. Instead of the goal conditions being the satisfaction of some representational conditions, the goal conditions are the satisfaction of some functional conditions — which don't have to be representational. The goal subsystem **A**, then, is a conditional switch: when its functional conditions are satisfied, it switches out of the **A-B** system; and, when they are not, it switches to **B**. Such a conditional switch, in turn, is simply a control flow process with two (or more) outcomes in which the process, thus the outcomes, is controlled by some other process or condition. If the controlling process yields a switch out of **A-B**, then such a process constitutes a satisfier of the goal conditions; if the controlling process yields a switch to **B**, then such a process does *not* constitute a satisfier of the goal conditions. In a classic trivial case, temperature controls the bending of a bimetallic strip in a thermostat (here the control is directly causal), and thereby either switches on a furnace or air conditioner, or switches off the system.

A Trouble with Functionalism. The framework outlined above is adequate for a partial development of the interactive model of representation, but it is not without its own potential problems. In particular, the notion of function itself can be problematic (Bechtel, 1986; Bigelow and Pargetter, 1987; Block, 1980b; Boorse, 1976; Cummins, 1975; Neander, 1991; Wimsatt, 1972, 1976; Wright, 1973). I will not address all of the several potential problems here, but there is one problem that is central not only to the idea of function per se, but even more so to the task of explicating emergent representation. To this problem, I will outline a solution.

Observer Functionalism. The notion of function was analyzed above in terms of consequence. The crucial problem has to do with the selection of *functionally relevant* consequences out of the multitudinous available causal relationships and consequences at the physical level of analysis of a system: Which consequences are relevant to the functional characterization of the system? In particular, if this problem of functional characterization is itself analytically arbitrary on the part of an observer and analyzer of the system, then we would have an observer functionalism, in which all functions are so only relative to the analytic choices of the observer. As is sometimes claimed, anything can be described as a Universal Turing Machine with the right descriptive choices. I'm not convinced that the issue is that wide open — descriptions *can* be false, and a UTm does require some (small) minimal complexity — but, nevertheless, the specter of an observer functionalism is quite real.

In particular, any explication of representation based on an observer functionalism will simply generate a more-complex-than-usual observer semantics — and we already have too many of those. In this sense, the general problem within functionalism of explicating function in a non-epiphenomenal and non-observer dependent fashion — of explicating functional analysis within naturalism — has a particularly strong relevance to the problem of representation. If the analysis of function is itself dependent on the intentionality of an observer, then to explicate the intentionality of the 'aboutness' of representation in terms of function will be viciously circular. I wish to indicate, then, how observer-dependent functionalism can be avoided; that is, how functionalism can be explicated naturalistically.

Normativity. One source of the problem involved in doing this is that the idea of function involves a normative aspect — a function is something that is *supposed* to be served, and a subsystem can succeed or fail in doing so — and naturalizing normativity is intrinsically difficult. What is needed is a sort of criteria for functional success and failure that is not itself already intentional.

Epiphenomenality. Intentionality could be avoided in defining criteria of functional success and failure simply by removing the observer from consideration, and taking into account only the factual matter of whether or not particular criteria have been satisfied. This may remove intentionality at least one step, if not eliminate it, but it introduces its own serious problem: epiphenomenality. There may be a fact of the matter concerning whether or not a system satisfies some arbitrary criteria, but if that satisfaction or lack of satisfaction makes no naturalistic difference to anything else in the universe, then the satisfaction of such a criterion is causally and ontologically epiphenomenal. It is irrelevant to the nature and functioning of the world, and, therefore, to understanding the world. Intentionality creeps back in at this point in that the only way in which the satisfaction or lack of satisfaction of such criteria can have a non-epiphenomenal effect in the world is via the mediation of the intentionality of an observer that can note and respond to that satisfaction or lack thereof.

Certainly some functional analyses have exactly this sort of status. An esthetic functional analysis of the parts of a painting, for example, would be purely epiphenomenal if not for the responses of intentional evaluators of such esthetics. Similar points hold for the functions of tools and other designed, or at least used or intended to be used, systems. In fact, a great deal of functional analysis does ground in the intentionality of observers and users in this manner. There is not necessarily a problematic in this, not even a problematic for naturalism, so long as the intentionality itself can ultimately be made good within a naturalism. But, if it is (some aspect of) intentionality itself that is at issue and to be modeled, and, more specifically, the naturalism of intentionality, then this is precisely the vicious circularity that must be blocked or avoided.

Functional Explanation. Another approach to the problem of functionalism is in terms of functional explanation: the existence of **A** is explained in terms of the function it serves for **B**. If **B** is itself an intentional agent, then we simply have an observer or user or designer functionalism. If **A** and **B** are parts of a biological system, however, then appeal can be made to the evolutionary history of systems of that type, to the species history, in such explanation. In general, the functions that **A** serves for the organism as a whole are arguably not epiphenomenal so long as those functions have contributed to the reproductive success of that species in the past. Such historical contributions to the satisfaction of evolutionary selection pressures, in turn, are taken as at least partially explanatory of the existence of such subsystems, such organizations of system processes, in contemporary organisms.

Functional explanation in this sense can be viewed as an aspect of evolutionary theory as a whole, and some version of this story must be at least part of functionalism more broadly within biology. As a general solution to the problem of functionalism, however, this approach suffers serious limitations. A crucial problematic for my purposes is the notion of function for systems that do not have such evolutionary history. Without such a history of contributions to meeting selection pressures, this approach leaves little ground for functional analysis. This is both a conceptual problem, in that the possibility of an animal coming into existence purely by accident would not seem to preclude a functional analysis of its organs and organ systems, and a regress problem, in that it grounds function in current systems in the functional history of systems of that type, of that species. Such a historical regress must halt. There must be some earliest system for which functions have a non-epiphenomenal reality *for the first time*, and this emergence must still be accounted for. A second perspective on this same point is that such an historical approach does not address (though it might arguably provide grounds for addressing) the problem of normativity. The normativity of functional analyses applies to contemporary instances of systems, and, at least in the case of designed and other intentional systems, does not require an evolutionary history.[5]

Process Ontology. There are several aspects to the framework within which I propose a naturalistic model of function. The ground for all of them is a process ontology. The development of a process ontology is itself potentially problematic (Birrell and Davies, 1982; Brown and Harré, 1988; Kitchener, 1988; Lucas, 1989; Shimony, 1986), and I do not find any contemporary versions to be satisfactory, but I will set aside that level of concern for the purposes of this discussion.

Emergence. Within a process ontology, phenomena of *emergence* are, in principle, not problematic. There is no particular naturalistic mystery about new organizations of process being able to instantiate new *properties*. The properties emergent in the organization of a computer constitute a motivating example, as do the properties of water that are not sums or aggregations of properties of hydrogen and oxygen. Similarly, if *everything* is constituted as organizations of underlying processes, then the emergence of new *ontological types* in new organizations, new patterns, of process is, in principle, not naturalistically problematic. The familiar complex hierarchy of nucleons, atoms, (stars, galaxies), molecules, cells, organisms, (ecosystems, biospheres), minds, societies, and so on, is a hierarchy of such process patterns emergent in underlying process patterns.

Organization. The perspective I propose focuses centrally on process *organizations*. At a given level of emergence, there are only two ontologically relevant aspects: (the properties of) those organizations of process available — ones that have already emerged and currently exist — and the new organizations or patterns of those available suborganizations that might come into existence. Organization, in this view, is central to emergence, and, thus, to virtually all ontology, at least above quantum fields (and arguably at that level too). Organization, then, is a core of a process metaphysics — at least of the one that I propose.

It is well understood that organization makes a difference, but organizational issues are generally set aside as initial or boundary conditions in explanations of particular phenomena, and not understood as metaphysically central. Typical metaphysics of substance and property (e.g., Kim, 1991) strongly motivate such a neglect of organization in ontological considerations. Furthermore, not only are there motivational misdirections from substance and property metaphysics, such metaphysics create serious conceptual and logical difficulties, if not impossibilities, in even attempting to address such *relational* phenomena as organizations of process (Olson, 1987). Simply, organizational relations are neither substances nor properties, and are not ultimately explicable in terms of them.

Ontological Epiphenomenality. New patterns of process, then, can instantiate emergent new properties. These too could simply be epiphenomenal, but, if those new properties have *causal consequences*, then, in that sense, they are causally *not* epiphenomenal. Even a process pattern that is not *causally* epiphenomenal, however, could nevertheless be, in an important sense, *ontologically* epiphenomenal. This would be the case, for example, if the coming-into-existence of instances of such patterns is purely accidental, and if their existence once instantiated is fleeting. In such cases, the pattern and its emergent properties may play a role in explaining and understanding ensuing phenomena, but it serves the conceptual function of initial or boundary conditions, not that of part of the "furniture of the world." The ontological reality of process patterns, then, depends in some sense on the continued existence of instances of those patterns.

There is much more conceptual exploration to be done here, but the main conclusion that I wish to draw for this discussion of functionality is that the ontologically strongest version of emergence in organizations of process is one in which among the emergent properties is that of the stability or persistence of the pattern itself. In such cases, the historical fact of the initial coming-into-existence of the pattern instances not only yields the causal consequences of those particular instances, it also changes the ontology of the world via the introduction of those patterns and their properties. Correspondingly, it changes what can and does happen, including further possible emergences, in that world.

Downward Causation. The consequences of such emergent ontological introductions can permeate all levels of ontology, including levels *below* that at which the emergence occurs. In other words, such emergences can be not-epiphenomenal not only at their own and higher levels, but at all levels. The reality of such emergences in the sense of consequence can, in principle, be found at all levels. One example of such "downward causation" is provided by soldier termites in certain species (D. Campbell, 1974a). The jaws of these termites are so large that the individuals cannot feed themselves, and depend totally on being fed by other members of the society. Such jaw size is adaptive for the colony for the specialized function of defense which these soldiers serve. But the explanation of the existence of such jaws, and, therefore, of the existence of such arrangements of proteins and other molecules — *issues at a level of analysis far below that of the termite society and species itself* — requires the emergent properties of the evolutionary process in general and of those of the social character of the species in particular. Ontological emergence is consequentially real.

Creation and Stability. There are two senses of the ontologically real emergence of process patterns. They both involve the increased probability of instances of the pattern in the future. One sort of increased probability derives from a given instance increasing the probability that new instances will come into being. Examples would be auto-catalysis and circular catalysis (Jantsch, 1980) and reproduction. The second version is simply the tendency for the persistence of a given pattern instance, once any such instance has come into existence. Note that evolutionary processes involve an interaction of reproductive pattern continuation and individual instance stability.

Energy Well and Open System Stabilities. The stability of process pattern *instances*, in turn, has two fundamental sub-versions. The first is one in which the pattern is stable so long as external impacts sufficient to disrupt the pattern do not occur. These are generally 'energy well' stabilities, in which energy is required to disrupt the pattern, and, so long as that threshold of energy is not available, an instance of the process pattern will persist. Examples of such energy well stabilities are nucleons, atoms, and molecules. The second sort of pattern instance stability is constituted by processes that require continued transactions with their environments in order to exist — open systems. These too can be disrupted by above-threshold inputs, but, unlike the first case, they *require* appropriate levels and kinds of input and output transactions with their environments in order to exist at all. Examples would include flames and living things.

Naturalistic Functions. Given this framework of process organizations and their emergents, and of the ontological reality of emergents grounded in pattern persistences once instances of those patterns are created, accounting for the naturalistic character and emergence of 'function' is relatively straightforward. Broadly, naturalistic criteria for functions derive from contributions to ontological emergence. That is, if a consequence of a (sub)pattern is to increase the probability of the existence of that pattern, either the same instance or new instances, in the future, then that consequence is ontologically functional by virtue of, and in the service of, that increased probability. Most specifically, such consequences are neither intentionally dependent nor epiphenomenal. This analysis is applicable even to such cases as auto-catalytic molecules, though usually relatively trivially. More interesting cases involve contributions to the stability of single process pattern instances, and paradigm cases — normal biological cases — make contributions in both the reproductive and the single instance stability senses.

Roughly, then, the ground that I propose for the modeling of naturalistic function is in two parts. First, process organizations, including functional organizations and control organizations, can instantiate emergent properties — witness the properties emergent in the organization of a computer that are not instantiated in the simple aggregation of the parts. Second, such emergent properties will, in some cases, contribute to the probability of existence of instances of that process organization, either in the sense of contributing to the stability of a given instance, or contributing to the evocation or construction of new instances, or both. At this point, the emergent property is no longer merely in the analytic perspective of the observer. It now has consequences independent of the observer and outside of the level of analysis of the emergent property itself — namely, whatever physical (and other) consequences follow from the persistence of instances of such organizations as distinguished from their absence.[6]

Developing the ground. An emergent property or consequence is ontologically functional insofar as it contributes to the increased probability of instances in the future. The most salient cases for this discussion are those involving open systems. Within that category, paradigm cases involve the functional contribution of some subsystem for the overall system, and most commonly involve issues of evolution as well as of single system stability.

Self Maintaining Systems. The most primitive forms of functional emergence, however, do not involve evolution and do not involve part-whole differentiation into subsystems. As mentioned above, perhaps the most primitive cases involve auto-catalysis, but there are some interesting developments in primitive functional emergence within *open systems* that are especially relevant that I would like to briefly explore. The relevancies have to do not only with the *conceptual* ground for functional, and therefore representational, analyses, but also with the *etiological* ground for functional and representational systems. These developments involve emergent properties of the overall open system process that contribute to the stability of existence of that process. I call systems with such emergent contributions to their own stability *self-maintaining*.

A flame, for example, has a functional property of being (partially) self-maintaining in the sense that it contributes to the maintenance of one of its own existence conditions — high temperature. Such a self-maintaining organization can, if other conditions are 'right', be enormously consequential in its persistence and spread. Self-maintenance is itself, then, *already* an emergent functional property in a sense that is independent of any observer, and is non-epiphenomenal.

Recursive Self-Maintenance. Consider now a system that is not only self-maintaining, but is *recursively* self-maintaining in the sense that it tends to maintain its own property of being self-maintaining. In order for an instance of a *recursively* self-maintaining system to be non-trivial, there must be alternative possibilities — more than one — of self-maintenance processes among which some system process can select. Furthermore, since conditions for stability are relative to various aspects of an environment, in order for there to be any actual tendency for those selections to improve the self-maintenance of the system, those selections must be based on some control influence from the environment — some control influence that tends to activate self-maintaining processes in appropriate environmental circumstances. That is, there must be more than one manner in which the system can be self-maintaining, and it must be able to switch them on and off, or switch between them, depending on differentiating influences from the environment. Such an organization is not only functional, it is a primitive control organization.[7]

The significance of this point is that only with recursive self-maintenance — only with a selection or choosing of self-maintaining processes — will a system be capable of *adjusting* to environmental variation, as distinct from simply surviving or not within the various environmental conditions. What has emerged here, along with the primitive control organization, is a primitive version of *adaptivity*. A *simply* self-maintenant system **is** stable or is not stable across some range of environmental conditions — it is adapted to that range, or not. A *recursively* self-maintenant system can adjust so as to **become or remain** stable across some range of environmental conditions — it is adapt*ed* across some range for *each* of its available self-maintaining processes, and is adapt*ive* across the union of the conditions of its various adaptednesses that it can switch among.

Adaptive Evolution. If a recursively self-maintaining system is also self-reproducing, then variations in reproduction will introduce variations in recursive self-maintaining ability — in adaptiveness. Variations in adaptiveness could be constituted, for example, in variations in the ranges of environmental conditions to which the systems could adjust. Variations in adaptiveness, in turn, are available for *differential selection* by the encountered environments. Here, then, we have the emergence of the possibility for the evolutionary *improvement of adaptiveness* of system types.

Recursive switching among self-maintaining processes can get *better*, more adaptive, either by becoming more appropriately sensitive in the switching; or by having additional self-maintaining processes — appropriate in differing circumstances — to switch among; or by having more powerful self-maintaining processes to use, even for circumstances for which some such process is already available; or any combination thereof.

Correspondingly, the variations in recursive self-maintaining systems will be differentially competent to variations in and ranges of selection pressures. Some systems will be able to adapt to one class of environmental variations, while others will be competent to a different class of environmental variations. Furthermore, some will be "better" than others either in the relative sense of capturing competencies for more common environmental variations, or even in the strong sense of being competent for a superset of such variations relative to some other recursively self-maintaining system. Variations in adaptiveness, then, are not restricted to variations in adaptive *scope*, but can exhibit variations in adaptive *ability*. With the emergence of recursively self-maintaining systems, therefore, *variations* in self-maintaining ability will be manifested, *and thereby subject to and available to evolutionary selection pressure*. This grounds the macro-evolutionary emergence of increasing kinds and scopes of adaptiveness, which, I argue elsewhere, generates the emergence of higher forms of mental phenomena such as knowing, learning, emotions, and consciousness (Bickhard, 1973, 1980a; Campbell and Bickhard, 1986).

Both Etiology and Analysis. For the current discussion, however, the important points are: 1) *Simple* self-maintenance is a primitive form of open system *functional* ontological emergence that can ground both further conceptual analysis *and* further evolutionary elaboration, differentiation, specialization, and emergence; 2) *Recursive* self-maintenance constitutes a primitive form of open system *control structure* ontological emergence that can ground both further conceptual analysis, and further evolution. In particular, this framework can ground both the conceptual analysis and the evolutionary emergence of interactive representation; it can ground solutions to both the analytic and the etiological concerns regarding representation. We will find, in fact, that even at this level, there is already a primitive version of interactive representation (see below).

So, at this point, we have *emergent* properties, emergent *functional* properties, and emergent *control* organizations that have non-epiphenomenal and non-intentional ontological reality, and that can conceptually ground the functional control structure analyses of interactive representation. And we have the emergence of primitive (functional, control structure) adaptive systems that can ground analyses of the evolutionary origins of such systems. With this framework, we can proceed to interactive representation itself.[8]

Interactive Representation. Assuming now a functional control flow framework of analysis, I turn to an explication of representation. Consider a (sub)system, with some particular internal control structure, in interaction with its environment. The course of internal activities in the system will depend jointly on what that internal control organization is and on what inputs are being received from the environment. (Note that those inputs may themselves be being induced by the outputs from the system.) In particular, when the interaction is completed, the (sub)system will end in some one of its internal states — some one of its possible final states. Some environments will leave the system in that same final state, when interactions with this (sub)system are complete, and some environments will leave the (sub)system in different possible final states.[9]

Implicit Definitions and Differentiations. The final state that the system ends up in, then, serves to implicitly categorize together that class of environments that would yield that final state if interacted with. A possible final state, then, implicitly defines, in an interactive sense, its class of environments. Dually, the set of possible final states serves to differentiate the class of possible environments into those categories that are implicitly defined by the particular final states. The overall (sub)system, with its possible final states, therefore, functions as a *differentiator* of environments, with the final states implicitly defining the differentiation categories.

These notions are generalizations of ideas already in the literature. Interactive implicit definition, for example, is an interactive generalization of the sense in which a formal language implicitly defines its class of models (Quine, 1966).[10] A differentiator is an interactive generalization of an automata-theoretic recognizer (Eilenberg, 1974; Ginzburg, 1968; Hopcroft and Ullman, 1979).

Such differentiators, and their environmental differentiations, are critically important emergents, *but they are not representations*. They implicitly define, they open-endedly differentiate, environments, but they do not represent anything about those environments. They do not constitute or carry any sort of representational content. Differentiators constitute an important aspect of the model of representation that I am developing, but they are not sufficient for full representation, and they are not even necessary to representational content.

Encoding Interpretations of Differentiators. This point deserves further attention, because it is precisely at this point that the typical encoding model goes awry. Differentiations of environments, within a class of possible such differentiations, constitute precisely the correspondences — and covariations of those environmental correspondences — that are so commonly taken to be encodings.

An observer of such a system and its environments could *characterize* the implicitly defined class of environments that are differentiated together by some final state — perhaps they all share some observer noticeable property — note the *correspondence* between such characterized environments and the given final state, and the *covariations* of similar characterizations for the other differentiated categories and their correspondent final states, and conclude that the given final state constitutes an *encoding* of its corresponding environmental characterization. This is precisely the form of standard, observer dependent, correspondence/covariation-as-encoding models. Typical examples of such purported correspondences-as-encodings — sensory 'transducers', connectionist networks, and so on — are just passive (no interacting outputs), and usually relatively simple, versions of an interactive differentiator. Standard models, then, construe differentiators with no representation whatsoever concerning what is being differentiated as encoding representations of the instances and categories being differentiated. But they have no representational content, and, therefore, cannot be encodings. Factual correspondences created by differentiators will turn out to be very useful, even representationally useful, for systems — and the factuality of such correspondences will help explain that functional usefulness — but those correspondences are not themselves adequate to representation.

Interactive Functional Predications. Consider now a differentiator in a broader control organization context — specifically, in the context of a goal directed organization. To make discussion simple, I will assume that the differentiator has only two possible final states, and, thus, two differentiation categories, **A** and **B**. I will also assume that the goal has two available possible interactive strategies, **S120** and **S137**, with their own corresponding final outcome states. The critical control organization for the emergence of representation is an environmentally sensitive differential switch in the context of a goal.[11] In this simplified case, we may assume, for example, that if the differentiator has arrived at final state **A**, then the goal system invokes strategy **S120**, for the sake of its internal outcomes, while if the differentiator has arrived at final state **B**, then the goal system invokes strategy **S137**, for the sake of its internal outcomes. (Note that the outcomes of the two strategies may, in successful executions, be the same; the strategies may differ with respect to the environments in which they are capable of inducing those internal outcomes.)

The first essential aspect of this is that, within such an organization, final state **A** functionally indicates the potentiality of **S120**, and final state **B** functionally indicates the potentiality of **S137**. If the goal is itself activated — if the goal exerts control on system activities — then these indications are followed. Such indications constitute functional implicit predications. In particular, in this example, **A**-type environments are indicated to also be **S120**-type environments, and **B**-type environments are indicated to also be **S137**-type environments. That is, "**A**-type environments are **S120**-type environments" and "**B**-type environments are **S137**-type environments." These predications could be wrong. If an **A**-type environment yields activation of **S120**, and **S120** fails — fails to achieve any of its internal outcome states, perhaps enters an infinite loop, or encounters an undefined condition — then the predication has encountered a counterexample, and it cannot be generally true. (Note that there is the possibility of another kind of error, an instrumental error, in this organization: an interactive strategy could succeed in producing one of its outcome states, but switching to that strategy could nevertheless still be a functional error if those outcome states do not induce satisfaction of the switching conditions for the goal.)

Error of the System, by the System, and for the System. Furthermore, the failure of such a strategy, and the falsification of the predication that is constituted by the functional indication of that strategy, is detectable and functionally available to the system itself: such failure *is* a functional condition in the system itself. If the system were to have some additional process to engage in upon such failure — perhaps some sort of variation and selection learning system, for example (Bickhard, 1980a, 1992a; Campbell and Bickhard, 1986; D. Campbell, 1974b) — it would have functionally available the relevant functional conditions to trigger that auxiliary system.

The claim is that such differentiated functional indications in the context of a goal directed system constitute representation — emergent representation. The indications predicate that differentiated environments have interactive properties appropriate to the indicated strategies. That is, they predicate interactive properties of those differentiated environments. Furthermore, those predications can be in error, and can be functionally detected to be in error from within the goal-directed system itself.

There are three critical aspects: 1) the environmental differentiations provide epistemic contact with the world — they differentiate environments, and then, on the basis of those environmental differentiations, indicate which interactive predications are available in those environments, 2) the indications themselves constitute implicit predications of interactive environmental properties — they constitute the representational content predicated of the differentiated environments, and 3) the embedding of such predications of representational content in the context of a goal directed system makes the falseness of those predications functionally detectable within the system, by the system, and for the system. This is representation, with content, capable of error, for the system — and therefore *not* observer dependent — that is emergent in functional control structure organizations that are themselves not already representational. This is genuinely emergent representation.[12,13]

Questions about Interactive Representation. There are many questions to be addressed about such interactive representation, most of which will not be discussed here.

There are two, however, to which I would like to at least indicate the nature of the answers.

What about Objects and the Rest of Our World? The first has to do with the fact that representational content, as explicated, is of environmental interactive properties. The question concerns the relationship between this form of representational content and the more familiar world of objects located in space and time, with causal connections, and so on. The general form of the answer involves two parts.

First, representational content is constituted as indications of potential further interactions. A given environmental differentiation might indicate not just the single strategy in the simplified example, but might indicate myriads of possible further interactions. Such indications will not, in any complex epistemic agent, be exhausted by single differential switches, but will comprise potentially vast complex webs of such indications. Furthermore, the indications of a particular environmental differentiation will in general not have the context-independent character of the example. What the final state of some environmentally differentiating interaction indicates about further potential interactions may depend not only on that interactive outcome, but also on the indications of interactive potentiality in many other domains of the web — on indications based on many prior interaction outcomes. Single environmental differentiating interactions, then, do not so much ground the *construction* of the web of indications of interactive potentialities as they ground *apperceptive updates* of that web (and such updating processes form the core of a non-encoding model of perception, Bickhard and Richie, 1983). Such apperceptive updates, their processes, and their properties, are a major domain of development within the overall interactive model, but simply pointing out the potential complexities of such an organization of indications is all that is needed here (Bickhard, 1980b; Bickhard and Richie, 1983).

The second part of addressing the question concerning representations of objects in space and time, and so on, is to note that certain patterns of interactive indications in the overall web will manifest interesting and potentially useful invariance properties. For example, some sets of potential interactions reciprocally indicate each other in the sense that the potentiality of any member of the set inclusively indicates the potentiality of the rest of the set. A visual scan of a manipulable object indicates in this reciprocal manner the potentiality of all the other potential visual scans of that object linked by appropriate manipulations of the object to bring those other aspects into view. Furthermore, that entire pattern of interactive potentialities remains invariantly available — linked by appropriate manipulations, translations, locomotions, and so on — with respect to a large class of other potential interactions, of movements and manipulations of that object. That is, physical manipulable objects afford a pattern of reciprocally indicative interactive potentialities that remains invariant under manipulations, translations of the object, locomotions of the agent, and many others. It is not invariant under such interactions as setting afire.

The basic proposal is that objects — from an epistemic perspective, not a theoretical or metaphysical perspective — *are* such invariant patterns of indications of interactive potentiality. Each infant spends major portions of the first two years of life constructing and elaborating knowledge of such pattern invariants and their interrelationships (Piaget, 1954). As adults, we may reflect on such invariances and inquire concerning their explanation, generate such theoretical notions as molecules, atoms, and so on, but the original epistemological characterization is in terms of such invariants. A note about their usefulness: such invariants permit the system to update its web of indications of potential interactions without having explicit current interactive grounding. So long as an invariant pattern was indicated, and so long as no indications that would destroy such a pattern have occurred, then the pattern should remain available, perhaps linked to the system's current situation by various intermediate interactions. You assume that your living room is still interactively accessible, for example, even though it may require a plane trip, taxi ride, commuter train ride, and so on to access it — unless you obtain information to the contrary. Invariant patterns, in other words, expand our worlds beyond immediately available perceptual environments. Objects then, are epistemological invariants that are representationally useful precisely because of such invariance properties: it is in terms of such invariances that our world is extendable beyond immediate interactive access. Locations in space and time, casual relationships, and so on are, epistemologically, related complicated representational constructions and developments (Bickhard, 1980b; Piaget, 1954).

What About Things like Numbers and Other Abstractions? The second question concerning interactive representation that I would like to briefly address concerns its adequacy to abstract representation. Even if it is granted that interactive representation might be adequate to representations of physical environments, it might seem impossible for this model to explicate representations of abstractions, such as numbers — where are the abstract environments of numbers to be interacted with? There are many properties of interactive representations that differ from those of the standard encodings of contemporary Cognitive Science, several of which are relevant to this point. I wish to introduce only one of those properties here.

The simple answer to the question is to point out that such an abstract environment does in fact exist, and exists potentially to be interacted with. In particular, the properties of the interactive systems and of their interactive processes are themselves more abstract than that which they interact with. Such properties could be interactively differentiated and represented within a higher level system interacting with the first level system that interacts with the environment. Such a second level system, in turn, would manifest properties that might be useful and could be represented from a third level perspective. And so on. The interactive model, then, generates a hierarchy of levels of potential knowledge, each representing properties of the level below it, with the first level interacting with the external environment. Instead of being discomfited by the problem of abstract representation, then, the model offers an unboundedly rich approach to it. Furthermore, this is not an ad hoc model purely to account for abstract representation — it generates a model of cognitive development that has its own independent implications and support in the literature (Campbell and Bickhard, 1986).[14]

Two prima facie challenges to the interactive model — that of objects and that of abstract knowledge — turn out to have independently motivated answers. This discussion has at best been generally indicative of the nature of those answers, and it has only addressed two such possible challenges: it is intended to be only illustrative of the claim that the interactive model might be competent to the multitudinous manifestations and forms of representational phenomena. I haven't addressed, for example, learning and development (Campbell and Bickhard, 1986) or language (Bickhard, 1980b, 1987; Bickhard and Campbell, 1992) or perception (Bickhard and Richie, 1983) or concepts (Campbell and Bickhard, in preparation) or rationality (Bickhard, 1991a), and so on.

Solving and Dissolving the Challenges to Encodingism. There are many directions of development of the interactive model. But one major domain of questions are those that have been addressed to and within the encoding framework. I would like to show how interactivism avoids those problems — especially the incoherence problem.

Interactivism is not Encodingism. First, notice that an interactive differentiation is not an encoding: it carries no representational content. Furthermore, interactive representational content is also not an encoding: the content is emergent in indications of interactive potentialities, and is not borrowed from, nor a stand-in for, anything else. Still further, interactive representations can ground the definition of derivative encodings: it is perfectly possible to define a stand-in for an indication of some sub-web of the web of interactive indications. Under certain circumstances, it will in fact be highly useful for a system to develop such internal secondary encodings (Bickhard and Richie, 1983), but they remain always subsidiary.

Emergence, Coherence, Construction, and Error Checking. Interactive representation and representational content are emergent out of non-representational phenomena. They do not encounter the aporia of an atomistic substance metaphysics trying to account for its own atoms. Since the representational content of interactive representations emerges out of system functional organization, it does not encounter the encodingist incoherence of having to borrow that content from itself. Error in interactive representation is internally detectable and potentially correctable in a quasi-evolutionary variation and selection constructive process,[15] so it does not require that the system already know the world in order to construct a copy of it. And interactive representations are constituted in organizations of indications of interactive potentialities, so when an interactive implicit predication is falsified, that falsification is itself emergent in the processes of the system. There is no possibility of questioning the interpretations or correctness of the encodings that separate the system from its world in this model, because the epistemic contact is directly constituted in the interactions with that world, not mediated by a veil of encodings. Error, then, when it occurs, is inherently constituted in system processes, and is therefore not subject to skeptical challenges — however true it also is that the *reasons* for error and the nature of possible *corrections* are at best defeasibly discoverable. Thus, since the world is interactively implicitly defined (though not arbitrarily or freely so), not unknowably corresponded to, and since those implicit definitions are emergent in interactive system organization, error is internally emergent and functionally available, not hidden behind an impenetrable veil of encoding correspondences.

Interactivism, then, is not subject to the fundamental aporias *of* encodingism: the impossibility of emergence, the incoherence of content, the circularity of construction, and the impossibility of checking. Furthermore, it is also not subject to the standard problematics *within* contemporary encodingism. Error, for example, is difficult even to define within a correspondence-as-encoding framework, while the connections between environmental differentiations and indicated further interactive potentialities are purely contingent — the possibility of error is trivially accounted for within the interactive model. The problem of correspondences with too many things is similarly irrelevant to the interactive model — what is being represented is directly emergent in the representational content in the interactive model. Neither the representational predications nor the representational contents predicated are correspondences in this model, so the multitude of factual correspondences are of no relevance to what is being represented. The differentiations do construct internal states that will be in factual correspondence with many things and conditions in the environment, but all that is required in the interactive perspective is that one of those correspondences ground the functional *usefulness or appropriateness* of the representational content being ascribed, not that one of those correspondences provide that representational content itself. It does not matter epistemically, for example, that activities in the retina may be in factual correspondence with many many conditions in the light, electron orbitals, and so on, so long as the ascribed potentiality of the interaction of "walking on a surface" in fact holds.

Such factual correspondences may play important roles in explaining why the ascription of such interactive potentialities, such representational contents, tend to be useful to the system, and, in that sense, can be quite important to an observer analyzing the system, but they do not constitute nor provide any of the representational content itself.[16]

Again, these worries in the contemporary literature are pure red herrings. They are the manifestations of confusions introduced by the incoherencies of the encodingisms that dominate that literature.

On What Hasn't Been Addressed. The interactive model as presented here has attempted to characterize only the most minimal property of genuine representation. In particular, it attempts to model the emergence of *internal* functional information concerning the falsification, or lack of falsification, of something that has an "aboutness" concerning the environment. Specifically, the falsification, or lack of falsification, of indications of the potentiality for particular interactions with that environment. With respect to intentionality more broadly, and even more so with respect to mentality more broadly, this is indeed minimal. Much more needs to be constructed on this foundation in order to begin to fill out those broader concerns.

First Level — Simplification. In particular, there are at least three levels of consideration at which the presentation in this chapter is simplified or incomplete. The first level is a set of potential complexifications within the basic organization of interactive representation already outlined. A differentiator, to begin, can have more than two final states, and, in most real cases, will. The set of possible final states for a differentiator might have some structure on it, such as an ordering or a metric or something more complicated. The differentiating final states, for example, might be well-ordered analog signals from a measuring process. The final states might also have structure internal to each state. "Final states", for example, might themselves be constituted as organized patterns of activity in the overall system.

There might be multiple indications of interactive potentialities from a given final state. The switch enabling aspect of indication could hold with many many other switching processes, goal processes, in the system. The outcome of a visual scan of a glass, for example, might indicate both the possibility of drinking and the possibility of throwing. Such indications might be context dependent, perhaps very complexly context dependent, on other final states, and on other indicators that have themselves been set on the basis of prior indicative context dependencies. The construction of interactive indications, and *webs* of such indications, might, in fact, be enormously complicated, and constitute an important internal system dynamic of its own — *apperception* (Bickhard, 1980b; Bickhard and Richie, 1983). The goals involved might be part of complex hierarchies of goals and servomechanisms.

Second Level — Simplification. The second major level at which the model presented thus far is simplified is a representational level version of the problem of causal epiphenomenality discussed with respect to functional analysis. In particular, although a system of the sort modeled will generate internal functional information concerning the falsification, or lack thereof, of its interactive indications, very little will follow from, will be consequential on, such falsification or its absence. As discussed thus far, the only consequence is whether the subsystem switches out of itself, or back to its strategies — and that differential switching **is** the functional information of success or failure of the previous strategy interaction.

Representational Epiphenomenality. There is a sense, then, in which, although functional analysis per se has been rescued from causal epiphenomenality (or so I claim), the representational systems as analyzed to this point are themselves *representationally* epiphenomenal, or, perhaps, their representationality is itself causally epiphenomenal — little depends on or follows from the fact that such systems minimally realize emergent possibilities of falsifications of "aboutness".

Trivial is Good. There are two responses to this recognition: First, in general, any model that does *not* have trivial or epiphenomenal versions is not likely to have evolved. If the simplest version of a system in which some property or phenomena is emergent is too complicated, then that emergence will be unlikely to occur, unless by preadapted accident, so that it can be responsive to selection pressures for further elaboration and improvement. If it *can* occur trivially, however, then it is much more likely to come into existence at all, and then be developed further in evolution. In this sense, the triviality or epiphenomenality of the minimal model is a virtue: it's possible for even one celled life to realize such trivial control structures.

Non-trivial is Easy. Second, it is not difficult in principle to recognize further system processes that could depend crucially on such internal information concerning interactive success and failure, thus rendering the interactive representation not epiphenomenal. Hierarchies of servomechanisms, for example, with switching consequences across the hierarchy depending on such success and failure begins to capture such non-epiphenomenality. But a quite strong consequence would be the invocation by interactive failure of a *meta*system that engages in trial constructions of new system organization, but remains generally inactive under conditions of interactive success. Such a metasystem will stabilize only with system organizations that yield interactive success. Under appropriate conditions, it will constitute the emergence of a minimal *learning* system, and, in this system context, interactive representational success or failure will be quite consequential (Bickhard, 1973, 1980a). The critical point, then, is that in such minimal systems the functional information *is (emergently) available* for the system, or for evolution, to do further things with, even if little or nothing further is done with it in the minimal systems per se.

Third Level — Incompleteness. A third level of consideration at which the model presented is simplified and incomplete is with respect to the multiple unaddressed properties and phenomena of intentionality and mentality, especially human mentality. These include perception, memory, learning, emotions, consciousness, development, language, the self, values, rationality, personality, psychopathology, and so on. I have addressed each of these at least briefly elsewhere, and some quite extensively (e.g., Bickhard, 1973, 1980b, 1987, 1989, 1991a, 1992a, 1992b; Bickhard and Campbell, in preparation; Bickhard and Christopher, in press; Bickhard and Richie, 1983; Campbell and Bickhard, 1986, 1992, in preparation). The general project, however, is clearly open ended, and not destined for any sort of completion. The role of interactive representation within this broader project is to serve as a ubiquitous principle of organization upon which and within which other intentional and mental processes can be modeled.

Location in a Broader Project. There is a general characteristic of this project, however, that I would like to comment on. I view these many mental and human phenomena as having evolved progressively over the course of macro-evolution. There are serious constraints on the sequence in which they could have evolved, but they did not spring into existence all together (Bickhard, 1973, 1980a). The significance of that obvious point is that, if that is so, then the gulf between mental and non-mental cannot be the singular void that Descartes leaves us with, even after rejecting his dualistic account of that void. If such a model of progressive evolutionary emergence is correct, then there are many mental properties and kinds of mental processes, and some of them can exist without others, while some of them require the prior existence of others in order to function or to have evolved themselves. Instead of a void, then, this picture is of complicated potential evolutionary trajectories, involving the sequential and progressive evolution of more sophisticated mental emergents, and differentiations and elaborations of prior emergents. The Cartesian void is filled with a rich structure of trajectories of emergence; The Cartesian diremption of mental from non-mental is healed.

In such a model of progressive emergence, questions of the demarcation of mental from non-mental take on a different form. In particular, it becomes arbitrary where to draw a line below which mind does not exist, and above which it does. Instead, the evolutionary trajectories themselves, and the points of emergence along them — both initial trivial points and later complicated elaborations — become the focus of interest. Mentality becomes a direction or trajectory, perhaps even tendency, of the cosmological evolution of the universe — a strong sort of naturalism. The minimal model of interactive representation presented here will be valid insofar as it can successfully participate in the construction of such broader models.

Fodor on Transduction and on Narrow Content: Two examples of encodingism confusions.

With an inspissation of the interactive model now at hand, I turn again to some of the issues that exercise contemporary encodingism. The perspective provided by interactive representation can illuminate even more deeply some of the confusions of encodingism: having an alternative perspective can help notice and diagnose errors that might otherwise go unnoticed because of underlying, implicit, shared encodingist presuppositions — after all, "What else is there besides encodings?"[17] Conversely, such analyses illustrate some of the properties of interactive representation and its relationships to standard conceptions.

Transduction. Fodor (1986), in addressing the question of whether or not his perspective commits him to the conclusion that paramecia have mental representations, claims that paramecia do not have mental representations because, although they can transduce environmental properties and respond selectively to the products of those transductions, they cannot make inferences — they can only respond to nomic properties in the environment since transduction is intrinsically nomic. Paramecia cannot respond to nonnomic properties, such as that of being a crumpled shirt, because selective responding to such nonnomic properties requires inference on the basis of nomically transduced properties. In Fodor (1991), he withdraws the defining characterization of transduction as being intrinsically nomic in favor of transduction being non-inferential.[18]

There are at least two aspects of this issue, that hold regardless of which version of transduction is considered, that I would like to point out. In both cases, the presuppositions of encodingism, and the consequent failure to take into account interactive possibilities, have confused and distorted the issues.

The first aspect has to do with whether or not transductions in paramecia produce representations. If so, then Fodor is committed to paramecia having representations in spite of the presumed lack of inference in paramecia. If not, then Fodor is in trouble concerning human transduction. And transduction can't create representations anyway.

The second aspect concerns the presumption by Fodor of the exhaustiveness of a dichotomy that underlies his arguments: the assumption that representations must be produced in one of only two possible ways, either 1) directly by transduction, or 2) mediatedly via inference. This pair of possibilities is *not* exhaustive, and, again, transduction cannot generate representations anyway.

First Aspect: Paramecia Transductions. First, the critique of encodingism lands directly on Fodor's notion of transduction — whether construed nomically *or* noninferentially — and the incoherence of the notion of transduced encodings shows up here in a confusion concerning transduction in paramecia and in humans. Transduction, in Fodor's framework, *must* yield representations, encodings in fact, in order for his account of human representation to work. Transduction provides the grounds, the material, for inference (when inference does or can occur, such as in humans), and inference cannot take place on just any naturalistic phenomena. Inference requires that representations be generated on the basis of prior representations — of propositions, in fact (Fodor and Pylyshyn, 1981). For Fodor, then, transduction — whether nomic or just noninferential — *must* yield encodings. But that is impossible,[19] and, if it *were* correct, then paramecia *would* have representations via their transductions, regardless of their lack of inference.

Perhaps Fodor could claim that paramecia do have representations, but they are just not *mental* representations because of the lack of inference. This would construe "mental" as "engages in or participates in inference". Humans and paramecia would be equivalent, on this view, with respect to having representations — via transduction — but the representations in humans would be "mental" because humans "engage in inference" with those representations. But this would be just a word game concerning how "mental" is to be used. The fundamental question remains that of the presumed representations themselves, and whether or not paramecia have them, and this maneuver would retain the commitment to paramecia having representations.

A somewhat more interesting possibility arises from the claim that transductions in paramecia *don't* produce representations because they don't have to: since paramecia don't engage in inference, they don't need representations upon which to base those inferences. This possibility, however, creates its own fatal problems.

In this perspective, transduction by paramecia does not *have* to yield encodings, because all that paramecia do is respond selectively, and simple informational correspondence is sufficient for that. Human transduction, however, *is* required to yield encodings, because it must ground inference. The difference between the paramecia and human beings, then, is not just that humans have inferences on top of transductions, but that humans require a fundamentally different kind of transduction — transduction that yields encodings, transduction that yields representation, transduction that yields *mental* (even if "unconscious") representation.

But now we have a position that claims that humans have representations because they have transductions that generate them (and must, in order to ground inference), and paramecia don't have representations because they have transductions that don't generate them (and don't have to since paramecia don't infer). The issue of whether or not paramecia have representations, then, has devolved into the issue of whether or not paramecia transductions are like human transductions in producing representations.

Fodor doesn't want them to be alike, because he wants paramecia to not have representations. He can claim that paramecia don't *need* representations since they don't infer, but then he must face the question of what the difference is between human transduction and paramecia transduction. Presumably paramecia transduction is just as nomological as human transduction, and presumably it is just as non-inferential too.[20] So, by either of Fodor's criteria, we have paramecia on par with humans concerning transductions. The lack of *ensuing* inference based on the products of paramecia transductions has no bearing whatsoever on the nature of what those transductions produce. There would seem to be little ground left for Fodor's claim that his position does not commit to paramecia having representations.

Fodor's claimed differentiation between humans and paramecia with regard to mental representations, then, turns out to be circular: humans have mental representations because they must (transductions must provide them, since humans engage in inference, and inference requires representation), and paramecia don't because they don't need them (transductions don't have to provide them, since paramecia don't engage in inference). That is, humans have (transduced) mental representations and paramecia have (transduced) non-representations, which is exactly what was supposed to be explained, not presupposed, in the first place.

Furthermore, the circularity of this argument is just a reflection of the circularity of encodingism in general. Real transduction can produce at best correspondence. That is all that paramecia need, so Fodor can leave them without representation. Humans, on the other hand, must have grounds for their inference, so transduction must produce more than just correspondence — transduction must produce representation, encoded representation. Human transduction *must* produce representation, in this view, so it is claimed that it does. But the grounds for that claim, the additions to correspondence that are supposed to turn transduction correspondence-products into representation — nomologicalness or non-inferentiality — apply equally to humans *and* to paramecia. So, either way, there are no non-question begging, non-circular, grounds for differentiating them.

If Fodor were to claim that transductions in paramecia *do* yield encodings, but that paramecia simply don't go on to engage in inferences based upon them, then 1) his distinction between paramecia and humans reduces to a distinction between no inference and inference, rather than no representation and representation, since both paramecia and humans are regarded as having transduced mental representations, and nobody is surprised or rhetorically impressed that humans and paramecia differ in that way, and 2) the mystery of where the transduced representational content comes from is simply pushed down to paramecia instead of the miracle occurring somewhere higher on the phylogenetic scale.

Second Aspect: The Transduction-Inference Dichotomy. The second aspect of this overall argument that I would like to focus on is that the exhaustiveness of Fodor's dichotomy between transduction and inference only holds from within his own framework of encodingist presuppositions. If encodings are to be generated, then that must either be on the basis of prior encodings, i.e., inference, or direct, i.e., transduction — whether nomic or not. But this assumes that the exhaustive basic issue is the generation of *encodings*, whether directly or inferentially, and *that* holds only if *all* representation is in fact encoding representation. The interactive model falsifies that encodingist assumption, and falsifies the assumed exhaustiveness of the transduction-inference dichotomy as sources of representation that is based on it.

An interactive differentiator does not yield an encoding, nor any representational content at all, yet it does ground functionally implicit predications, which constitute functionally implicit *inference*, but strictly in a sense that is implicit in the functional organization of the system, not in a manipulations-of-encodings or symbols sense. The inference here is strictly functional and strictly implicit; it is not syntactic nor formal; it is not Fodor's inference — it is not the generation of new symbol strings on the basis of already extant symbol strings.

So, where do such differentiators fit in Fodor's "exhaustive" dichotomy? Differentiators are not necessarily nomic (lawful), so they are *not* transducers according to Fodor's earlier characterization. But they can ground representations, and even derivative encodings, and, therefore, even Fodor type inference. So, by this criterion, they *must be* transducers — contradiction. Furthermore, if they are not transducers, then, according to Fodor's dichotomy, they could not ground any sort of inference. But if they are transducers, then, according to Fodor's dichotomy, they must yield encoded representational content — and they don't. Differentiators do not fit anywhere within this version of Fodor's classifications: they must be transducers and not transducers simultaneously.

On the other hand, with respect to Fodor's later transducer criterion, non-inferentiality, differentiators must be transducers: differentiators are not (necessarily) inferential — however much they may ground functionally implicit predicational inference, they do not engage in inference, and certainly not in Fodor's formal symbol manipulation inference, themselves. Therefore, by this characterization, differentiators *must* be transducers. Furthermore, differentiators can ground inference, so by this criterion too, differentiators are again transducers. But, for humans at least, differentiators-as-transducers — Fodor-transducers, anyway — must yield encodings as the ground for formal inference, and differentiators don't do that. At best, differentiators are non-encoding paramecia-type transducers, not Fodor-transducers, yet, unlike transducers in paramecia, differentiators ground representation. Again, according to this version, differentiators must both be and not be transducers. At a minimum, then, Fodor's exhaustive dichotomy between transduction and inference is a false dichotomy, and all of his reasoning based on it is invalid.

Interactive differentiation is not (necessarily) nomic; it is not (necessarily) inferential; it does not yield any representational content in itself, and, therefore, cannot yield encodings; yet it does ground the functionally implicit predication of interactive properties, and, thereby, the emergence of (at least potentially mental) representational content. As such, it is neither fish nor fowl within Fodor's dichotomy, and therefore destroys (the exhaustiveness of) that dichotomy.

Fodor presupposes that the only way to get *encoding* representational content is either direct or inferential, which is falsified by interactively based, interactively derivative, encodings; and he assumes that the only way to get representational content at all is — in one of these two ways — to get encodings, which is falsified by the interactive model in general. Fodor presupposes that representational content *is* encoding content.

So, interactive differentiators obliterate the coherence of Fodor's transducer characterizations: differentiators are not nomological (therefore, they are not transducers); they are non-inferential (therefore, they are transducers); they can ground inference, even (indirectly) symbol manipulation inference (therefore, they are transducers); but they do not (directly) yield encodings (therefore, they are not transducers). And interactivism in general destroys the exhaustiveness of Fodor's dichotomy of how to get representations — either via transduced encodings or via inferred encodings: interactive implicit predications of interactive potentiality are neither transduced encodings nor inferred encodings. They are not encodings at all.

More generally, the interactive model shows how representational content can emerge on the basis of non-representational differentiation — non-encoding, paramecia-type, transduction, differentiation — and, thus, it directly addresses the origin of the representational content that Fodor's story leaves a miracle. Finally, I reiterate the point that Fodor's argument against the Gibsonian position is based essentially on the presumption that this dichotomy is exhaustive, so that argument fails along with its presumption.

The most important point here, however, is not the failure of the arguments per se, nor the specifics of the confusions that underlie those arguments, but that those confusions are themselves products of struggling with and within the traps of the encoding perspective. The presumed dichotomy between inference and transduction, the confusion concerning whether or not paramecia 'transductions' yield representations, the confusion in the very question concerning the representationality of paramecia 'transductions' — all of these are themselves products of taking the encodingist perspective.

Broad and Narrow Content. Another issue that has arisen from within the correspondence-as-encoding framework constitutes a confused partial convergence with the interactive model. This is the issue of broad and narrow content (e.g., Fodor, 1987a, 1991). The issue arises from asking what it is, or could be, inside the epistemic agent that determines what the agent-to-world correspondence(s) are with — which, in this framework, amounts to asking what inside the agent determines encoding representational content.

In particular, it appears that external, or broad, content — what the correspondences are with — cannot be completely determined from within the agent. A major source of this intuition derives from Twin Earth scenarios: consider a Twin Earth that is exactly like this earth, in all details including the presumably naturalistic mental states of the inhabitants, except that on Twin Earth there is no H_2O, but instead there is some compound XYZ that plays the same roles as water on this earth. The point, then, is that identical mental states in twin epistemic agents — one here and one on Twin Earth — will correspond to H_2O here but to XYZ there. The conclusion is that mental states cannot completely determine what their own correspondences will be with — which, again, in this framework, constitutes not being able to determine what their own representational contents are.

Narrow Content. One move — Fodor's move — is to postulate narrow (representational) contents in the epistemic agent that, together with particular contexts, jointly determine the external correspondences — the broad contents. This makes use of a notion of a map from context to content, introduced by Kaplan (1979a, 1979b, 1989) for demonstratives: a word like "this" picks out one thing in one context of usage, and something else in other contexts. In effect, narrow content proposes that all concepts have this context-dependent-content character to some degree, though some concepts like demonstratives and indexicals will be extremely context sensitive, while others, presumably, will require contextual variations on the level of that from earth to Twin Earth in order for their context dependencies to show up.

Narrow content, however, proves to be a peculiar notion. It seems, for example, to not be specifiable at all, since any content specification would necessarily be in terms of other concepts, which would exhibit their own context dependencies, and, therefore, any purported *narrow* content specification would itself simply *exhibit* this context-to-content functional relationship rather than explicating it. Trying to sort out the complexities of, and the controversies over, narrow content has become a significant subtheme in contemporary encodingism (Loewer and Rey, 1991).

Partial Convergence with Interactivism. The partial convergence with the interactive model in all this is that an interactive differentiator, to a first approximation, does exactly what narrow content is supposed to do: create a context sensitive covariational correspondence between an internal condition and an external condition in a given contextual environment. In this manner, it creates epistemic contact with the world, and serves as one ground for the indicated relevance of representations in general. The problem of context dependency, then, has forced Fodor, and others, to propose something with functional properties very convergent with those of interactive differentiators.[21]

Divergences. Beyond this, however, divergences abound. First, Kaplan's proposal was for a context dependency of words, not of mental representations. From an encoding perspective, this is of relatively little matter since words are encodings of mental contents, which, in turn, are encodings of the world — and the encoding stand-in relationship is transitive. Issues of context sensitivity, however, become noticeably more complex in the interactive model of language, since words cannot be encodings of mental contents. As a first approximation, words evoke context sensitive differentiations of internal context sensitive differentiators — there is a double layering of differing kinds of context sensitivity (Bickhard, 1980b; Bickhard and Campbell, 1992).

Setting language considerations aside, however, we find now familiar confusions in the differences between narrow content notions and interactive differentiators. Differentiators do compute context sensitive correspondences with the world, as contextually open differentiations of the world, but there is no confusion that those correspondences, context sensitive or not, constitute representational content for the system itself. An encodingist recognition of essential context sensitivity is here still laboring under the basic encodingism correspondence-as-encoding confusion. Differentiators create correspondences-as-information, but not correspondences-as-representation, and there is no knowledge of what the correspondence is with at all, either in the *computation* of the correspondence itself nor in the *existence* of the correspondence itself once computed. Differentiators create informational relationships with the world, where information is in the control sense, not any representational sense, and what is controlled by that information is indications of further interactive potentialities.

Such (webs of) indications of further interactive potentialities are what constitute representational content in the interactive model, not the environmental differentiations that context sensitively *control* those contents. The core function of environmentally interactive differentiators is precisely to *control* the evocation of (indications of) interactive representational contents in a context sensitive manner, not to *constitute* those representational contents. Only with (appropriate) such environmentally context-sensitive controls of interactive representational content can the system as a whole remain functional — alive.

Interactive representational contents *are* (organizations of) indications of further interactive potentialities, and a viable system *must* manage to manifest appropriate environmental sensitivity of those indications, and, thus, appropriate environmental sensitivity of the actual further system interactions. Such "appropriate" environmental sensitivity of consequent ultimate system interactions *is* adaptiveness. Differentiators provide the sensitivity; correct implicit predications in functional indications of interactive potentialities provide (part of) the appropriateness; and executions of those indicated subsystem organizations (strategies) provide the interactions. And, overall, such functional organization constitutes emergent representation.

Encodings are Defined by their Contents. The narrow content-broad content issue is caught within still another of the paradoxes of encodingism. Representation is context sensitive, and must be context sensitive, in order for epistemic agents to remain stable, but encodings, in the standard view, are intrinsically *not* context sensitive. Encodings are carriers of representational content, and representational content is supposed to specify *what* the encoding represents. It makes sense within standard encodingism to postulate context sensitive *constructions* of encodings, but not context sensitive encodings themselves. To postulate context sensitive encodings per se seems simply confused, in the encodingist view, in the sense that whatever representational content gets context sensitively selected — by whatever process — "ought" to be the content of the encoding, *not* the process of selection nor the mapping that the selection process computes. That is, classic encodingism presses for equating the representational content of an encoding with the broad content that is context sensitively selected — with the other end of the correspondence — not with the context sensitive map from context to "content" that picks out that broad, external "content".

Any such selected broad content itself encounters still another variant of the circularities. If narrow content selections of broad content are of things or conditions actually in the environment, then we are again committed to a semantics strictly from the observer's standpoint, and have not even begun to address the issue of representation-for-the-system itself, since we still have no model that even attempts to explicate the system's content. While, on the other hand, if narrow content selections are of contents strictly *within* the system — *observer independent* system content, *as are the indications in the interactive model* — then those selected *internal* contents need to *already be there* in order to be selected, and we need a completely different account of what it is that is being selected, of what content-for-system really is, of what "broad" content really is, and encodingism does not and cannot address that without vitiating itself on the incoherence problem. It's no wonder that the notion of narrow content has been perplexing.

Jerry-rigging Again. Once again, the interactive model *intrinsically* exhibits what has had to be Jerry-rigged into the encoding perspective — in this case, the context sensitivities that are involved in representation. And the narrow content—broad content attempt to account for such context dependencies doesn't work anyway, since it is still committed to an observer semantics. Standard encodings are encodings, and, therefore, are representations, by virtue of what they are *supposed* to represent — by virtue of the representational content that they are presumed to carry. So encodingism encounters the circularity of requiring that the representational content be known in order for the representational encoding to exist, since something is an encoding only by virtue of carrying such content — but, if all representations *are* encodings, as encodingisms presume, then we are back at the necessity to already have representation in order to get representation. Introduction of context sensitivities into what an encoding represents — into the encoding stand-in relationship — complexifies the issues, and makes it more difficult to try to define an encoding in terms of what it is supposed to represent, more difficult to entertain the simple confusion that an atomic encoding *could* be defined in terms of what it is supposed to represent, but it does nothing toward breaking out of the basic circle. It does nothing toward accounting for the emergence of representational content out of non-representational phenomena.

In other words, narrow content still leaves representational content an unaddressed miracle. It is still an encodingism, and is still incoherently unable to account for its own presuppositions — to account for the representational contents that are necessary in order for the atomic level encodings to exist at all, whether context sensitive or not. The interactive model does explicate such representational emergence, and does so in a way that intrinsically manifests the context sensitivities that have driven the distinction between narrow and broad content.[22]

Conclusion

The primary task of this chapter has been to argue the impossibility of standard conceptions of, and approaches to, representation. The alternative interactive model has been developed primarily in order to demonstrate that the family of critiques of encodingism does not lead to an unescapable impasse regarding representation. I have shown that interactivism doesn't fall to the same impossibilities as encodingism, but there are many issues remaining that are crucial to demonstrating the adequacy of interactivism to representational phenomena in general. These include standard cognitive phenomena, such as perception, language, conceptualization, rationality, creativity, and so on, and further aspects of intentionality, such as the attitudinal aspects of propositional attitudes, the systematicity of propositional attitudes, emotions, and consciousness (Bickhard, 1980a; Campbell and Bickhard, 1986, in preparation).

Nevertheless, what has been developed in this chapter is already sufficient to show how representation can emerge out of non-representational phenomena, and, thus, how it can exist at all. It shows how representation could emerge and how it could evolve in a manner fully consistent with naturalism. It avoids the absurdities of radical innatism (Bickhard, 1991a, 1991b; Fodor, 1975, 1981), of correspondences-as-encodings, and the incoherencies of encodingist representational content.

In particular, the interactivist model of representation is fully capable of being instantiated in human beings and other animals and of being constructed in machines. Appropriate functional control structure organizations for representational emergence are no more mysterious or impossible for machines than for human, or other animal, nervous systems — whether innately wired or constructed.

Interactivism, however, does imply that no passive system can have genuine representation: representational content is constituted as indications of potential interactions, and that is not possible (at least not for the system itself) if the system is itself not capable of interaction. Artificial intelligence, insofar as it attempts artificial epistemic agents, must be a branch of robotics, not of computer science or programming theory. Similarly, cognitive science must, in this view, abandon the traditional information processing backbone of perception, to cognition, to language in favor of very different frameworks of analysis. For example, in the interactive view, perception is not an input phase or stage, but is instead "merely" a specialized kind of interaction — generally just like all other kinds of interactions (Bickhard and Richie, 1983). The study of language, similarly, takes on a completely different form within a framework that eschews utterances as encodings of mental contents (Bickhard, 1980b, 1987; Bickhard and Campbell, 1992).

Minimally, interactivism and its associated critiques of encodingism show that there are foundational flaws in contemporary approaches to representation, intentionality, and naturalism — flaws that arguably have been with us for millennia. More generally, interactivism shows how to avoid the multiple aporias of encodingism, but it requires a major rethinking of cognitive phenomena in return. It offers the possibility of a deeper understanding of, and a more powerful conceptual framework for, human cognition; and it offers the possibility of genuine artificial epistemic agents (Bickhard, 1991c).

References

Ashby, W. R. (1960). *Design for a Brain*. London: Chapman and Hall.

Bechtel, W. (1986). Teleological functional analyses and the hierarchical organization of nature. In N. Rescher (Ed.) *Current Issues in Teleology*. Landham, MD: University Press of America, 26-48.

Bickhard, M. H. (1973). A Model of Developmental and Psychological Processes. Ph. D. Dissertation, University of Chicago.

Bickhard, M. H. (1980a). A Model of Developmental and Psychological Processes. *Genetic Psychology Monographs, 102*, 61-116.

Bickhard, M. H. (1980b). *Cognition, Convention,and Communication*. New York: Praeger.

Bickhard, M. H. (1987). The Social Nature of the Functional Nature of Language. In M. Hickmann (Ed.) *Social and Functional Approaches to Language and Thought* (pp. 39-65). New York: Academic.

Bickhard, M. H. (1988). Piaget on Variation and Selection Models: Structuralism, Logical Necessity, and Interactivism *Human Development, 31*, 274-312.

Bickhard, M. H. (1989). The Nature of Psychopathology. In L. Simek-Downing (Ed.) *International Psychotherapy: Theories, Research, and Cross-Cultural Implications*. (115-140). New York: Praeger.

Bickhard, M. H. (1991a). A Pre-Logical Model of Rationality. In L. Steffe (Ed.) *Epistemological Foundations of Mathematical Experience* New York: Springer-Verlag.

Bickhard, M. H. (1991b). Homuncular Innatism is Incoherent: A reply to Jackendoff. *The Genetic Epistemologist*, *19*(3), p. 5.

Bickhard, M. H. (1991c). How to Build a Machine with Emergent Representational Content. *CogSci News*, *4*(1), 1-8.

Bickhard, M. H. (1991d). The Import of Fodor's Anticonstructivist Arguments. In L. Steffe (Ed.) *Epistemological Foundations of Mathematical Experience*. New York: Springer-Verlag.

Bickhard, M. H. (1992a). How does the Environment Affect the Person? In L. T. Winegar, J. Valsiner (Eds.) *Children's Development within Social Contexts: Metatheoretical, Theoretical and Methodological Issues*. (pp. 63-92). Hillsdale, NJ: Erlbaum.

Bickhard, M. H. (1992b). Scaffolding and Self Scaffolding: Central Aspects of Development. In L. T. Winegar, J. Valsiner (Eds.) *Children's Development within Social Contexts: Metatheoretical, Theoretical and Methodological Issues*. (pp. 33-52) Hillsdale, NJ: Erlbaum.

Bickhard, M. H., Campbell, R. L. (1989). Interactivism and Genetic Epistemology. *Archives de Psychologie, 57*(221), 99-121.

Bickhard, M. H., Campbell, R. L. (in preparation). *Topologies of Leaning and Development*.

Bickhard, M. H., Campbell, R. L. (1992). Some Foundational Questions Concerning Language Studies: With a Focus on Categorial Grammars and Model Theoretic Possible Worlds Semantics. *Journal of Pragmatics*, 17(5/6), 401-433.

Bickhard, M. H., Christopher, J. C. (under review). The Influence of Early Experience on Personality Development.

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. (in preparation). *The Impasse of Artificial Intelligence and Cognitive Science*.

Bigelow, J., Pargetter, R. (1987). Functions. *Journal of Philosophy, 84*, 181-196.

Birrell, N. D., Davies, P. C. W. (1982). *Quantum Fields in Curved Space*. Cambridge.

Block, N. (1980a). Introduction: What is functionalism? In N. Block (Ed.), *Readings in philosophy and psychology* (Vol. I). (171-184). Cambridge: Harvard.

Block, N. (1980b). Troubles with functionalism. In N. Block (Ed.), *Readings in philosophy and psychology* (Vol. I). Cambridge: Harvard.

Bogdan, R. (1988a). Information and Semantic Cognition: An Ontological Account. *Mind and Language, 3*(2), 81-122.

Bogdan, R. (1988b). Mental Attitudes and Common Sense Psychology. *Nous, 22*(3), 369-398.

Bogdan, R. (1989). What do we need concepts for? *Mind and Language, 4*(1,2), 17-23.

Boolos, G. S., Jeffrey, R. C. (1974). *Computability and Logic*. Cambridge.

Boorse, C. (1976). Wright on Functions. *Philosophical Review, 85*, 70-86.

Brown, H. R., Harré, R. (1988). *Philosophical Foundations of Quantum Field Theory*. Oxford.

Burnyeat, M. (1983). *The Skeptical Tradition*. Berkeley: University of California Press.

Campbell, D. T. (1974a). 'Downward Causation' in Hierarchically Organized Biological Systems. In F. J. Ayala, T. Dobzhansky (Eds.) *Studies in the Philosophy of Biology*. (179-186). Berkeley, CA: University of California Press.

Campbell, D. T. (1974b). Evolutionary Epistemology. In P. A. Schilpp (Ed.) *The Philosophy of Karl Popper*. (413-463). LaSalle, IL: Open Court.

Campbell, R. L., Bickhard, M. H. (1986). *Knowing Levels and Developmental Stages.* Basel: Karger.

Campbell, R. L., Bickhard, M. H. (1987). A Deconstruction of Fodor's Anticonstructivism *Human Development, 30*(1), 48-59.

Campbell, R. L., Bickhard, M. H. (in preparation). *Knowing Levels and the Development of Natural Kind Categories: Interactivism, structuralism, and nativism.*

Campbell, R. L., Bickhard, M. H. (1992). Types of Constraints on Development: An Interactivist Approach. *Developmental Review*, 12(3), 311-338.

Carlson, N. R. (1986). *Physiology of Behavior*. Boston: Allyn and Bacon.

Churchland, P. M. (1984). *Matter and Consciousness*. MIT.

Cummins, R. (1975). Functional Analysis. *Journal of Philosophy, 72*, 741-764.

Eilenberg, S. (1974). *Automata, Languages, and Machines. Vol. A* New York: Academic.

Fodor, J. A. (1975). *The Language of Thought*. New York: Crowell.

Fodor, J. A. (1981). The present status of the innateness controversy. In J. Fodor, *RePresentations* . Cambridge: MIT Press (pp. 257-316).

Fodor, J. A. (1986). Why Paramecia don't have Mental Representations. In P. A. French, T. E. Uehling, H. K. Wettstein (Eds.) *Midwest Studies in Philosophy X: Studies in the Philosophy of Mind*. U. of Minnesota Press, 3-23.

Fodor, J. A. (1987a). *Psychosemantics.* Cambridge, MA: MIT Press.

Fodor, J. A. (1987b). A Situated Grandmother? *Mind and Language, 2*, 64-81.

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.

Fodor, J. A., & Pylyshyn, Z. (1981). How direct is visual perception?: Some reflections on Gibson's ecological approach. *Cognition*, *9*, 139-196.

Fodor, J. A., & Pylyshyn, Z. (1988). Connectionism and Cognitive Architecture: A critical analysis. *Cognition, 28*(1-2), 3-71.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.

Gibson, J. J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting and knowing*. (67-82). Hillsdale, N.J.: Erlbaum.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Ginzburg, A. (1968). *Algebraic Theory of Automata*. New York: Academic.

Harnad, S. (1990). The symbol grounding problem. *Physica D*.

Hopcroft, J. E., Ullman, J. D. (1979). *Introduction to Automata Theory, Languages, and Computation*. Reading, MA: Addison-Wesley.

Horgan, T., Tienson, J. (Eds.) (1988). *Connectionism and the Philosophy of Mind. Southern Journal of Philosophy, XXVI, Supplement*: Spindel Conference, 1987.

Jantsch, E. (1980). *The Self-Organizing Universe*. Pergamon.

Kaplan, D. (1979a). On the logic of demonstratives. In P. French, T. Uehling, Jr., & H. Wettstein (Eds.), *Contemporary Perspectives in the Philosophy of Language*. Minneapolis: U. of Minnesota Press, 401-412.

Kaplan, D. (1979b). Dthat. In P. French, T. Uehling, Jr., & H. Wettstein (Eds.), *Contemporary Perspectives in the Philosophy of Language*. Minneapolis: U. of Minnesota Press, pp. 383-400.

Kaplan, D. (1989). Demonstratives: an essay on semantics, logic, metaphysics, and epistemology of demonstratives and other indexicals. In J. Allmog, J. Perry, H. Wettstein (Eds.) Themes from Kaplan. Oxford University Press, 481-563.

Keisler, H. J. (1977). Fundamentals of Model Theory. In Barwise, J. *Mathematical Logic*. North Holland.

Kim, Jaegwon (1991). Epiphenomenal and Supervenient Causation. In D. M. Rosenthal (Ed.) *The Nature of Mind.* Oxford, 257-265.

Kitchener, R. F. (1988). *The World View of Contemporary Physics*. SUNY.

Kneale, W., Kneale, M. (1986). *The Development of Logic*. Oxford:Clarendon.

Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). SOAR: An Architecture for General Intelligence. *Artificial Intelligence, 33*, 1-64.

Loewer, B., Rey, G. (1991). *Meaning in Mind: Fodor and his critics*. Oxford: Blackwell.

Lucas, G. R. (1989). *The Rehabilitation of Whitehead*. SUNY.

Lycan, W. G. (1990). The Continuity of Levels of Nature. In W. G. Lycan (Ed.) *Mind and Cognition*. Blackwell, 77-96.

McClelland, J. L., Rumelhart, D. E. (1986). *Parallel Distributed Processing. Vol. 2: Psychological and Biological Models*. MIT.

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the Structure of Behavior*. New York: Holt, Reinhart, and Winston.

Moore, G. H. (1988). The emergence of first order logic. In Aspray, W., Kitcher, P. *History and Philosophy of Modern Mathematics*. U. of Minnesota, 95-135.

Neander, K. (1991). Functions as Selected Effects: The Conceptual Analyst's Defense. *Philosophy of Science, 58*(2), 168-184.

Newell, A. (1980a). Physical Symbol Systems. *Cognitive Science, 4*, 135-183.

Newell, A. (1980b). Reasoning, Problem Solving, and Decision Processes: The Problem Space as a Fundamental Category. In R. Nickerson (Ed.) *Attention and Performance VIII*. Hillsdale, NJ: Erlbaum.

Nicolis, G., Prigogine, I. (1977). *Self-Organization in Nonequilibrium Systems*. New York: Wiley.

Nicolis, G., Prigogine, I. (1989). *Exploring Complexity*. New York: Freeman.

Olson, K. R. (1987). *An Essay on Facts*. Stanford, CA: Center for the Study of Language and Information.

Piaget, J. (1954). *The Construction of Reality in the Child*. New York: Basic.

Piaget, J. (1970). *Genetic Epistemology*. New York: Columbia.

Pippin, R. B. (1989). *Hegel's Idealism*. Cambridge U. Press.

Popkin, R. H. (1979). *The History of Scepticism*. Berkeley: University of California Press.

Prigogine, I. (1980). *From Being to Becoming*. San Francisco: Freeman.

Pylyshyn, Z. (1984). *Computation and Cognition*. MIT.

Quine, W V. (1966). Implicit Definition Sustained. In W. V. Quine *The Ways of Paradox and other essays*. (195-198). Random House.

Rumelhart, D. E., McClelland, J. L. (1986). *Parallel Distributed Processing. Vol. 1: Foundations*. MIT.

Shimony, A. (1986). Events and processes in the quantum world. In R. Penrose, C. J. Isham (Eds.) *Quantum Concepts in Space and Time*. (182-203). Oxford.

Smith, B. C. (1985). Prologue to "Reflections and Semantics in a Procedural Language" In R. J. Brachman, H. J. Levesque (Eds.) *Readings in Knowledge Representation*. (31-40). Los Altos, CA: Morgan Kaufmann.

Smith, B. C. (1987). *The Correspondence Continuum*. Stanford, CA: Center for the Study of Language and Information, CSLI-87-71.

Van Gulick, R. (1982). Mental Representation: A Functionalist View. *Pacific Philosophical Quarterly*, 3-20.

Waltz, D., Feldman, J. A. (1988). *Connectionist Models and Their Implications*. Norwood, NJ: Ablex.

Wimsatt, W. C. (1972). Teleology and the Logical Structure of Function Statements. *Studies in the History and Philosophy of Science, 3*, 1-80.

Wimsatt, W. C. (1976). Reductive Explanation: A functional account. In R. S. Cohen, C. A. Hooker, A. C. Michalos, J. Van Evra (Eds.) *PSA-1974. Boston Studies in the Philosophy of Science*. (Vol. 32, pp. 671-710). Dordrecht: Reidel.

Wright, L. (1973). Functions. *Philosophical Review, 82*, 139-168.

[1]. I have mentioned three stances that are commonly taken with respect to a potentially representational system: a designer stance, a user stance, and an observer stance. The three stances differ in terms of the *origins* of the knowledge of correspondences between conditions internal to the system and those external to the system: a designer typically stipulates and engineers such correspondences; a user typically learns them; and an observer typically diagnoses them from observations and analyses. What all three have in common, however, is that they constitute perspectives that are simultaneously on both the system *and* on that system's environment. It is by virtue of this dual perspective that encoding correspondences can be defined — between known conditions in the system and known conditions in the environment. *But these are known only to the holder of such an external perspective*. None of these three perspectives is that of the system itself, and the system *does not have* any independent perspective on its own environment. Such correspondences may (or may not) exist between the system and its environment, but to take any such correspondences as constituting representations for the system is to require that the system know not only the internal states, but also the correspondences and what those correspondences are with. Such a project encounters precisely the circularity of incoherence: the system must already know what the correspondences are with in order for the correspondences to constitute representations for the system, but such knowledge is what those correspondence representations were supposed to provide in the first place.

[2]. A camera produces correspondences, *lawful* correspondences, between the image at the back of the camera and scenes in front of the camera. Of course, no one confuses such images with representations — at least not for the camera itself. The imaging in a camera does not involve transduction in its typical sense, but the only function of the energy transformations of transduction in standard approaches is that it provides the grounds for the claims of lawfulness of the correspondences produced.

[3] In spite of this double recognition — 1) of the distinction between correspondence = correlation = covariation = information, on the one hand, and genuine representation, on the other, and 2) of the fact that the latter problem is still essentially untouched — we still find Fodor presuming, as if the matter is settled, the core of his claimed demolition of Gibson (Fodor and Pylyshyn, 1981; e.g., in Fodor, 1991). That critique of Gibson, however, is based essentially on a non-sequitur *equivocation* on exactly the distinction between information = correspondence and genuine representational content — from transduction, which gives at best correspondence, to "that the light is so-and-so" — representation. Bickhard and Richie (1983) show just how bad that "demolition" of Gibson really is.

Incidentally, we also find in Fodor (1991, p. 257) the assumption "that we have a story about representation along the lines of representation-is-information-plus-asymmetric-dependence", in spite of the apparent recognition of the distinction between information and representation.

In yet another instance of Fodor's attack on Gibson, we find in Fodor (1986, p. 19) the claim that "Gibsonians require that for *each* nonnomic property to which we can respond selectively, there must be a coextensive, transducer-detectable, psychophysical invariant; e.g., a light structure in the case of each such visual property." Gibsonians require no such thing. Fodor is again simply charging the Gibsonians with the consequences of his own non-sequitur. It is Fodor's confusion that yields the presumed exhaustive dichotomy between transduction and inference, and, therefore, it is Fodor's confusion that concludes that if the Gibsonians deny inference, they must be committed to transduction for everything. In fact, Fodor's transduction cannot exist (Bickhard and Richie, 1983) — it directly encounters the incoherence problem — and the dichotomy between transduction and inference is *not* exhaustive — it only even appears plausibly exhaustive, in fact, within a very narrowly focused encodingist perspective. Within such a narrow encodingist focus, you either generate encodings directly — transduction — or you generate them on the basis of earlier encodings — inference (see Fodor on Transduction in the main text below).

[4]. The notion of functionalism as involving the possibility of substituting for components or aspects in virtue of their (the substitutes) serving the same function as that substituted for does not endorse the Pylyshyn (1984) notion of a singular cleavage between biological and functional levels of consideration. In fact, the hierarchical process-emergence model within which this discussion proceeds forces, and, thus, converges with, something like Lycan's (1990) model of hierarchical functionalism.

[5] There is an approach to this regress problem that construes current functions of a subsystem type in terms of past contributions to meeting selection pressures in ancestral organisms, and holds that the *first* instances of such subsystems, presumably accidental, did *not* in fact serve a function. Functional emergence, then, depends on past contributions to the satisfaction of selection pressures; and past satisfaction of selection pressures as a ground for this emergence is taken to be at least one step closer to an ultimate naturalistic model. This maneuver, however, does not address the conceptual problem of systems without evolutionary history, and does not address the normativity problem.

It also relies on the choice of 'satisfying selection pressures' as somehow criterial for 'function', but does not address why the satisfaction of selection pressures should itself be taken as criterial for anything. The implicit response to this charge is that it is in terms of such selection pressure satisfaction that the existence of the current system is to be explained. I would argue that there is a germ of a solution hidden within this implicit answer, but, prima facie, it simply backs up the regress one step: Why should contributions to the existence of current systems be taken as criterial?

From my perspective, the deepest problem with this approach is that it fails to render function as non-intentional and non-epiphenomenal. Whether or not a system has a function in this view supposedly depends on whether or not there is an appropriate history of contributing to the satisfaction of selection pressures, but the presence or absence of such a history *makes no difference whatsoever to the current activities of the system, and to its consequences*. The presence or absence of such a history, and, therefore, in this account, the presence or absence of there being functions to be served at all, is causally and ontologically epiphenomenal. An intentional observer might conclude that there was or was not such a history, and, therefore, that there was or was not a function in this sense, but that does not provide ground for a naturalism of functional analysis.

[6] Note that since this analysis is in terms of *patterns or types* of process organizations, it is already intrinsically also in terms of *tendencies or propensities* for certain consequences to obtain. It thus addresses naturally such potentially problematic examples as malfunctioning instances of system components, or components that are never called on for the performance of their function in this particular system instance.

That is, it addresses normative issues in virtue of the modalities of propensity at the level of types. I will not address here the complexities that can obtain when, for example, normative functional analysis and functional etiological explanation interact, as for vestigial organs. Clearly, such interactions and differentiations can ramify with impressive complexity.

[7]. Self maintenance and recursive self-maintenance can be construed as functional propensities (Bigelow and Pargetter, 1987), except that here the propensity to contribute to the recursive self-maintenance of the system is taken to be a property of the *organization* of the underlying *process* in the system, not a property of some *part* of the system.

[8] Note that this framework relies on properties that were emphasized by the cybernetic ground out of which Artificial Intelligence and Cognitive Science developed, but which were abandoned in the shift to the symbol manipulation framework. The symbol manipulation framework as an approach to cognition, however, presupposed away one of its own fundamental problems — the nature of representation. In effect, I am proposing that the problem of emergent representation cannot be solved without taking into account several of the aspects of cybernetics that have been ignored — plus evolutionary considerations, a better grounding of functional analysis, and so on. On the other hand, it should also be noted that cybernetics itself had no satisfactory model of representation either.

[9]. The notion of a final state, and of a set of *possible* final states, is essentially that of the conditions in a system that are functional for exerting control on other parts of the system. The "final" part of the term is strengthened if the subsystem in question switches off only when in some final state, though little of the interactive model requires that. It would be quite in keeping with the model developed that a differentiator would function continuously and concurrently with other parts of the system, and would change its "final state" that exerted control influences on (indications for; see below) the rest of the system from time to time on the basis of its interactions. The term "final state" is adopted from automata theory, and certain aspects of automata theory, such as the discreteness and strict sequentiality, are not necessary for the interactive model. It is also possible, and, in fact, very interesting consequences follow, if a system's differentiators collectively can *store* their final states as indicators. For example, there might be a dynamics internal to the system that could ensue among those stored indicators, even though the *initial* indicators would be stored on the basis of interactions with the environment (Bickhard, 1980b; Bickhard and Richie, 1983).

[10]. This notion of implicit definition apparently originated in the 19th century with the realization that appropriate axiomatizations of geometry could be taken as implicitly defining the notions of geometry (Kneale and Kneale, 1986). It was adopted by Hilbert as a formalist approach to mathematics in general, in which all of mathematics would be so implicitly defined (Moore, 1988). It carries on in formal model theory in more refined notions such as those of categorical or monomorphic axioms (Kneale and Kneale, 1986) or of an elementary class of models (Keisler, 1977). It is this notion of axioms implicitly defining a class of models (Kneale and Kneale, 1986; Quine, 1966) that is generalized in the idea of interactive implicit definition.

There is also a related notion of the implicit definition of a *term* in an axiom system by its position and role within that system, and a proof that the possibility of such an implicit definition implies the possibility of (a somewhat ad hoc) *explicit* definition

(Boolos and Jeffrey, 1974; Kneale and Kneale, 1986; Quine, 1966). The interactive notion of implicit definition is *not* the definition of a term, but this theorem is nevertheless of relevance in showing that implicit definition is not intrinsically less powerful than explicit definition (Quine, 1966) — when explicit definition is possible at all.

It is this last point, of course, that is at the core of the issue: such explicit definition is possible only in terms of already available representations — it is an encoding definition — and, therefore, cannot serve any fundamental epistemological functions. Explicit definitions cannot yield new representational content or new knowledge; they can only yield stand-ins for representation and knowledge already available. Implicit definitions can.

[11] This notion of environmental sensitivity is with respect to the environment of the system under analysis. In the case of the central nervous system, for example, the 'environment' includes the rest of the body, and sensitivity to, say, blood sugar level, would constitute a functional sensitivity of an appropriate kind. The key point is that such sensitivity — to blood sugar level — is a strictly functional notion, and does not require any representation, of blood sugar or anything else.

[12]. Note also that while a system in which the function of representation is served must be a goal directed system — else there would be no internal criterion of error — the converse does not hold. Not all goal directed systems will be representational. In particular, only those goal directed systems that differentiate their activities in the service of the goal in accordance with differentiations of the environment (or some other source of such informational differentiation) will constitute minimal representational emergence according to this model.

[13] Note that, on this explication, any genuine adaptiveness will be at least primitively representational (such as a recursive self-maintenant system). This is in the sense of involving indications of interactive potentialities *about* the environment — such (indications of) interactive potentialities *are* the representational content.

Such a notion of content is more primitive than any involving systematicities or constituents of such content, or involving any differential attitudes towards (the predication of) such content. These issues, which commonly exercise Fodor and others, are changed in major ways by the interactive implicit definition explication of content, but will not be addressed in this chapter.

The general notion that representation must be functional for a system's interactions with its environment in order to functionally exist for the system at all is similar to Van Gulick (1982), but that explication proceeds, beyond the basic intuition, within a clear encodingist framework. There seems to be an assumption that only covariational correspondence *could* serve such an interactive function.

[14] With respect to number: one property of a lower level control system — or its executions — is the iteration of a subsystem within a larger system: a count of the number of iterations. This could be detected and controlled from within a given system level, but could be *represented* only from within a next higher knowing level.

[15]. Interactivism per se is a model of the nature of representation. In particular, it is not a model of learning or development. Interactivism, however, does have strong implications

for learning and development. In particular, interactivist system organizations cannot be passively impressed into a mind via induction or transduction, and, therefore, they must be actively constructed. Interactivism, then, forces a constructivism. Furthermore, such constructions, unless they are prescient, must be blind trials, so interactivist constructivism must be a blind variation and selection constructivism, an evolutionary epistemology (Campbell, 1974b). Variation and selection constructivism, in turn, is a kind of interactionism between system and environment with regard to learning and development. So, interac***tivism*** as a model of representation forces interac***tionism*** as a model of development. The two models of interactivism and interactionism, then, are not the same, not even models of the same phenomena, but they do have a strong relationship, with the first forcing the second (Bickhard, 1988, 1991d, 1992a, 1992b; Bickhard and Campbell, 1989; Campbell and Bickhard, 1986).

[16] For still another example, note that the interactive model of emergent representational content is not subject to the peculiar dissolution into "meaning holism" (Fodor, 1987a, 1990). The minimal interactive representation system, and its representational content, is far from holistic.

[17] A reaction I have received several times at talks that I have given on interactivism and the encodingism critique.

[18]. Note that by defining transducers as non-inferential rather than in terms of their being nomic, Fodor risks being committed to accepting connectionist nets as transducers. Among other consequences, he could no longer simply claim (as in his argument against his construal of Gibson, for example) that non-nomic properties, such as that of being a "crumpled shirt", cannot be "transduced."

[19]. To recapitulate, it is impossible for transductions to yield encodings because a transduced encoding must encode that which produces it, that which is energetically transduced, and, in order to do that, the system must already know what is being transduced — as usual, you must already have encoding representational content in order to get encoding representational content. The temptation to pass the origin of this knowledge onto prior learning or on to evolution fails because the same problem recurs at all levels: encoding representational content must come from somewhere, and neither evolution nor learning nor transduction can generate it (within the encodingist framework). In all cases, the question of the origin of the knowledge, the origin of the representational content, cannot be addressed. In all cases, the incoherence problem is encountered (Bickhard and Richie, 1983, contains a detailed tracking of the counters and counters-to-the-counters that might be proposed in this argument).

[20]. As mentioned before, so also is a camera.

[21] The Twin Earth intuitions turn on the presumption of naturalism: the conceptual experiment wouldn't work if thoughts were taken to be entertainments of Platonic forms. In particular, if naturalism is correct, then representation, whatever it is, must be constitutable in physical terms, or perhaps in physically instantiated functional terms. But functional interrelationships can at best functionally differentiate, differentiate with respect to functional properties — they do not provide a privileged epistemic access to *noumena*. So, if all physical relations are identical in two situations, then so also are all functional relations, and naturalism cannot epistemically penetrate deeper than to differentiate down to, but not below or within, functionally identical properties. Twin earth scenarios postulate a situation in which H_2O and XYZ have functionally identical

properties, even though they are in fact different compounds, and turns on the intuition that whatever is in the mind could not differentiate more finely than those interactive functional properties — without violating naturalism. Therefore, there must be something in the mind that picks out external differentiations, and such picking-out must be functionally context dependent. Assuming naturalism, then, whatever is representational in the mind cannot select or pick out or differentiate more finely than identity of interactive functional properties, and, therefore, such differentiation will necessarily be functionally context dependent.

If two substances were identical in functional properties with respect to *all possible* functional properties, then there are no grounds for supposing their difference at all. So, the Twin Earth intuitions also lead to the conclusion that if a party on either earth had access to both H_2O and XYZ it would be possible to find differentiating functional properties. Similarly, science on either world might have already found properties of H_2O or XYZ respectively that *would* differentiate it from the other should the opportunity to examine both occur.

Twin Earth arguments, then, turn on a basic principle of interactivism: the functional — but functional in the interactive sense, not in the Fodor encoding sense — character of representation, and the consequent impossibility for representational differentiations to partition more finely than those interactive functional properties. To attempt to capture those differentiating properties in terms of "contents" — narrow content — is to attempt to capture a differentiation process in terms of what it differentiates, but what is differentiated is specifiable only in a fully context dependent manner. But the differentiation functional organization itself is *not* context dependent, only its executions and its external conditions of outcomes are, and, therefore, once again, the encodingism approach presupposes itself. Once again, it must have content in order to get content, and it is impossible to have that content in the first place — no necessarily context *dependent* characterization of content, narrow or otherwise, can possibly capture the context *independent* functional organizations that context *dependently* differentiate such contents. So long as content is what is represented, and what is represented is what is corresponded to, the circularities of encodingism are unavoidable. *Real* narrow 'contents' — interactive differentiators — have, in themselves, no content at all. And they cannot be given content in terms of what they differentiate or correspond to. Among other consequences, they cannot be ultimately understood or explicated or modeled in terms of any such content.

There is still another encodingism-created problem that appears here: naturalism forces a distinction between narrow and broad content; folk psychology, meanwhile, seems to require that representational contents have causal interactions with each other, *and that these interactions be in terms of broad content*; yet only narrow contents are in the brain, and, therefore, only narrow contents are potentially causally available to each other. One move would be to attempt to demonstrate that folk psychology can get by with only causal interactions among narrow contents, but, however, superficially plausible this might seem in a Twin Earth case, its plausibility disappears for more sensitive Kaplanesque context sensitivities such as indexicals. A different move might be to deny folk psychology completely, but such denials must ultimately offer some sort of replacement (Churchland, 1984). The phenomena of talk, and of talk of beliefs and desires, and of causal consequences of such talk, may or may not end up being understood in standard folk psychology construals, but the phenomena must nevertheless be accounted for in some way. (The encodingism argument, in fact, has as a consequence that folk psychology in its standard interpretation *cannot* be correct — propositional attitudes— whether beliefs, desires, or any other sort — cannot be attitudes toward encoded propositions.)

This apparent conflict between naturalism and folk psychology, however, is itself a product of encodingism. Naturalism may force the narrow-broad content distinction, but resources for representational content, and for the potentiality of causal interaction among representational content, are *restricted* to broad or narrow contents only from within an encodingism.

In the interactive model, in particular, the representational content is *not* given in the context sensitive differentiator at all, nor in its "broad" differentiated externality, but, rather, in indications of potential further interactions — indications of the potentiality of executing other system organizations. The differentiators, the indications, and the other system organizations whose potential executions are indicated, are all aspects and parts of the system itself, and, thus, are all naturalistically available for possible interaction with each other *in that system*. Whatever may be the ultimate fate of folk psychology, it won't fail on the basis of the necessity of impossible broad-content causal interactions — at least not from within the interactive model. (There are other implications of the interactive model for folk psychology that I will not pursue here: one example is the sense in which an interactive system may function 'as if' it had certain beliefs — may function in ways that involve implicit functional presuppositions about the environment or the system itself, that do not involve *any* explicit representational content, Bickhard and Christopher, in press.)

[22] There is another interesting sense in which this entire standard debate is, from the interactive perspective, misguided from the start: the questions of intentionality and representation are being asked about "mental states". Interactivism is embedded in a strict process ontology, and, from such a perspective, to ask about mental states is equivalent to asking about flame states. Both flames and mental phenomena are intrinsically open systems, necessarily in process, and a state approach is at best an aspect of a mathematical idealization, and at worst a labyrinthian dead-end. Flames can have states in the sense of the idealization of a single time slice through the process, where the process itself is captured, say, by some differential equations, but there is no way to explicate what a flame is nor to understand it just in terms of such a state. Or a flame could have a state in terms of some relatively, even if momentarily, stable aspect of its process, such as "being in the state of burning that piece of wood". Again, however, although such "conditions of the process" may be quite relevant to some considerations, they cannot model what a flame is at all. There is no apriori reason why a state approach to mind should be any better suited than to flames, and the interactive perspective — for that matter, any reflection on the hierarchies of emergence of patterns of process through atoms, molecules, life, mind, and so on — argues strongly that a state approach is precisely such a labyrinthian blindness. Simply, a state approach is just an idealization of a substance approach. A state approach, a *mental* state approach, then, is bound to be fatally misleading as a presupposition of the very statement of the issues concerning mentality.