Troubles with Computationalism

Mark H. Bickhard

Mark H. Bickhard
Department of Psychology
17 Memorial Drive East
Lehigh University
Bethlehem, PA 18015
MHB0@LEHIGH.EDU

Bickhard, M. H. (forthcoming). Troubles with Computationalism. In R. Kitchener, W. O'Donohue (Eds.) *Psychology and Philosophy: Interdisciplinary Problems and Responses*.

Biographical Note

Mark H. Bickhard is a graduate of the University of Chicago, and was at the University of Texas at Austin from 1972 to 1990. He is currently the Henry R. Luce Professor of Cognitive Robotics and the Philosophy of Knowledge at Lehigh University. His primary interests concern the nature of psychological processes — including, among others, representation and language. Relevant publications include *Cognition, Convention, and Communication* (1980), *On the Nature of Representation* (1983), and *Foundational Issues in Artificial Intelligence and Cognitive Science: Impasse and Solution.* (1995).

Troubles with Computationalism

Mark H. Bickhard

Computationalism

Computationalism is the dominant contemporary approach to cognitive phenomena: phenomena of perception, cognition, reasoning, language — any mental phenomena that involve representation. Computationalism permeates the intertwined fields of Cognitive Science, Cognitive Psychology, and Artificial Intelligence. It grew out of cybernetics and computer studies during the 1950s and 60s.

Computers were originally thought of as very fast and powerful calculators. It came to be realized, however, that there is nothing in the functioning of a computer that restricted its domain to calculations and other manipulations of numbers. The electrical patterns that corresponded to numbers in computers could just as easily be taken to represent characters — or tables, chairs, propositions, perceptual features, grammatical structures, and so on, and on. During the 1960s, this move to a conception of computers as “symbol manipulators”, rather than “number crunchers”, became ascendant — and has remained so. The range of representation that has been thought to be capturable in such computer symbols has seemed limited only by the ingenuity of the designers.

The backbone of the computational approach to cognition is the presumption that symbolically encoded information flows from perception to cognition and from cognition to language. Perception is presumed to begin with the encoding of various sorts of limited and proximal information, such as light patterns and motions in the visual system. That initial limited information in proximal sensations is then enhanced in unconscious inferences about the organization of the layout of surfaces and edges — floors, walls, tables, etc. — and colors and so on (again in the visual system). Such perceptual encodings, in turn, serve as the foundation for the learning of symbolically encoded concepts and other more abstract representations, and serve to help guide the cognitive manipulation of such symbols in reasoning. Still further, such internal organizations and manipulations of symbols can themselves be re-encoded into linguistic form, and uttered or written for the recipients to decode into their own internal symbols. Decoded, they constitute an understanding of the utterances or sentences. Sensations of color, form, and shape, for example, might generate the perception of a lion, which, in turn, guides reasoning to the conclusion that there is danger, and further initiates both the activity of running, and vocalizations that, among other things, re-encode the mental symbol for “lion” into sound patterns, to warn others of the danger. Connected to this cognitive backbone are many other phenomena, such as memory, motivation, emotions, and so on, all of which also involve representation, and all of which are assumed to therefore involve manipulations of encoded symbols.

Since the early 80s, a major rival has challenged computationalism in its classical form. This rival is known as connectionism, or Parallel Distributed Processing (PDP). Connectionist systems consist of multiple nodes, usually in layers, and usually with the first layer directly connected to an environment. The nodes are connected, from each node in one layer to nodes in the next layer, by weighted lines. Activation is received from the environment, and passed along the weighted connections to other nodes in accordance with the various weights. Each node, then, adjusts its own level of activation in accordance with the activations received along its incoming connections, adjusted by the weights on those connections, and in accordance with some internal function for collecting and combining those activations. Then each node, in the next step, transmits its own level of activation along its outgoing connections to the next layer of nodes, where the process repeats.

Connectionist nets can be trained by various alterations of the weights connecting node to node in response to feedback from sets of training instances. In particular, they can be trained to generate particular patterns of final layer activations in response to various categories of patterns of input activations. That is, connectionist nets can be trained to differentiate input patterns into categories. This capacity, and the ability to train this capacity, has generated enormous excitement in a large community of cognitive scientists. The final activation patterns are taken to be representations of the classes of input patterns that would generate them, and, on this interpretation, connectionist nets can learn new representations. There are many additional properties of these distributed patterns of activation that are proposed as advantages of the connectionist approach, and many variations on the basic theme of input patterns classified into final layer patterns. Some net designs, for example, don't simply settle into some final layer activation pattern, but, instead, move into some classifying trajectory or cycle of changes in activation patterns, perhaps shifting among them in response to subsequent inputs. Heated arguments have been waged concerning the relative advantages and disadvantages of the two approaches.

The differences between standard computationalism and connectionism are many and important for a variety of reasons. But with respect to the basic critique that will be made of computationalism, there is no significant difference between the two — they both make the same basic assumptions, and they are both vulnerable to the same criticisms.

Between the two of them, computationalism and connectionism dominate contemporary cognitive science and related disciplines in the 90s. In spite of that, I wish to argue that both computationalism and connectionism are in serious trouble — fatal trouble, in fact. The problem lies in a ubiquitous presupposition about the nature of representation — the nature of those representational elements (or activation patterns) — a presupposition that is at root logically incoherent. I will present a central part of this critique, and adumbrate an alternative approach to representation that avoids this incoherence. And I will point out how the basic argument against standard conceptions of representation holds just as strongly against connectionist and PDP approaches: for all their differences, the approaches hold precisely the same assumptions — incoherent assumptions — about the nature of representation.

Encodingist Approaches to Representation

Throughout history, representation has been assumed to be some sort of correspondence between a representing element or structure and that which is to be represented, and, crucially, the representational relationship is assumed to be constituted in those correspondence relationships. It is universally recognized that not just any correspondence will do. There are myriads of factual correspondences in the universe, and most of them are not representational. The focus, then, is on what *sort* of correspondences are representational.

Various problems with this approach have been uncovered, and more are discovered periodically, but the assumption that "correspondences" is the correct genus within which to differentiate representation is still ubiquitous (Anderson, 1983; Fodor, 1990; Newell, 1980; Palmer, 1978; Smith, 1987). Each new problem generates new activity to find a fix — a further or a different constraint on correspondence that will capture all and only the real representations.

I hold that some of these problems are simply unsolvable from within this approach — they are among the multifold manifestations of the underlying logical incoherence. Others of these problems are simply red herrings — they appear to be problems *only* because representation is being approached as a kind of correspondence, and they cease to be problematic on the alternative that I offer (Bickhard, 1993).

Encodings and Encodingism. There *is*, in fact, a class of genuine representations that are constituted as correspondences: encodings. It is, in part, this genuine subclass that has produced so much confusion about representation, and maintained the impression that all representation could be of the same sort — that all representation could be encoding correspondences. It is from this class of representations that I derive one of the general names for the approach that I wish to criticize. Approaches that presuppose that all representations are encodings, I call *encodingism*.

The arguments against encodingism form a rather large family, with some of them having ancient provenance, and new ones being discovered even quite recently. I will outline only a few of these arguments. One convenient entree into this family is via the characterization of genuine encodings, and the demonstration that *that* kind of representation *cannot* be the general form. It is precisely the presupposition that it *can* be the general form that yields the logical incoherence.

So, I begin with a clear and obvious case of encodings: Morse code. Morse consists of a set of stand-in relationships between sequences of dots and dashes, on the one hand, and characters and numerals, on the other hand. For example, "**. . .**" stands-in for the character "**S**". Such stand-ins are useful because, in this case, dots and dashes can be sent over telegraph wires, while characters and numerals cannot. The stand-ins of Morse code serve to change the form and the medium of the representational contents of characters and numerals so that things can be done with them, such as telegraph transmission, that could not be done otherwise. Similarly, bit pattern encodings allow myriads of extremely fast manipulations in computers.

Encoding stand-ins carry representational content — they carry the same content as that which they stand-in-for. The stand-in relationship, in fact, is a relationship of "borrowing" the representational content of whatever is being stood-in-for. So, "**. . .**" represents the same thing as, performs the same representational functions as, "**S**". Representational content is a functional notion — it is whatever serves the function of specifying what is supposed to be represented by some representation.

There are several essential aspects and prerequisites of such encodings that are relevant for this discussion. First, an encoding stand-in correspondence relationship must be defined. This requires that the element that is to be the stand-in and the element that is to be stood-in-for must be both specifiable — in this case, both "**. . .**" and "**S**". Second, in order for the stand-in to carry representational content, the stood-in-for must carry representational content: the encoding carries representational content only insofar as it borrows it from the stood-in-for.

It follows that such stand-in relationships cannot be defined with elements (properties, events, etc.) that are themselves unspecifiable, and that an encoding stand-in can only borrow representational content that is already available (from the stood-in-for).

In particular, encodings cannot create *new* representational content. They can only borrow, and combine, representational content that is already extant.

These prerequisites are not a problem for genuine encodings, such as Morse code, precisely because genuine encodings are explicitly defined in terms of stand-in correspondence relationships with representational elements that are already available. The prerequisites are a problem, however, for any assumption that encodings could serve basic *epistemological* functions, such as in perception or cognition: the fundamental epistemological problem is to create new knowledge, new representation, and encodings can only borrow representational content that is already there.

It might seem that there could be some way in which encodings could generate or constitute new representational content. In fact, if it is assumed that representation *must* be some form of encodings — what else is there? — then it appears that there *has* to be some way in which encodings can generate new representational content, and the problem seems to be one of discovering a model that can account for that. This is the standard framework of assumptions in cognitive science, artificial intelligence, philosophy of mind and language, and so on. It is the assumption underlying computationalism and connectionism alike. And most of the work that concerns itself with this level of problem at all is devoted to attempts to discover such a model. I argue that this task is impossible, and that assumptions or presuppositions that it could ever be completed are incoherent (Bickhard, 1992, 1993; Bickhard & Richie, 1983).

As outlined, an encoding can be defined in terms of already available representations with no particular logical difficulty. Furthermore, this can be iterated, yielding "**X**", say, defined in terms of "**Y**", which is defined in terms of "**Z**", and so on. But it can only be iterated a finite number of times: there must be some ground, some foundation, of representations in terms of which the combinatoric hierarchy of encoding definitions can begin. It is at the level of this ground that the incoherence emerges of assuming that all representation is encoding.

Incoherence. In particular, if it assumed that such foundational representations are themselves encodings, then there is no way for those encodings to have or to be given any representational content. There is no way, in other words, for those grounding encodings to be representations at all, and, therefore, certainly no way for them to be encodings. But, then, it is impossible for them to ground any hierarchy of encoding definitions, and the entire framework of encodingism collapses into incoherence.

If a purported foundational encoding is defined in terms of any other representation, then it is not foundational, contrary to assumption. But that leaves only that foundational element itself as a source of representational content, yielding something like " '**X**' stands-in-for '**X**' " or " '**X**' represents the same thing as '**X**' " for "**X**" that foundational element. This does not suffice to provide any representational content to "**X**", and, therefore, fails to render "**X**" an encoding. "**X**", therefore, cannot ground any further definitions of encodings, and the framework collapses.

This problem of providing the basic representational contents has not gone unnoticed in the literature. It presents under such terms as "the empty symbol problem", "the symbol grounding problem", and so on (Bickhard & Richie, 1983; Block, 1980; Harnad, 1987). It underlies other positions and arguments, such as the "solution" that proposes that all such grounding representational contents are innate (Fodor, 1981). In that regard, note that the problem is logical, not practical — evolution cannot solve it either. Nativism is not a solution (Bickhard, 1991; Campbell & Bickhard, 1987).

Note that, as a *logical* problem concerning the origin of representation, this makes representation, considered as encoding, impossible. There is no way for grounding encodings to emerge out of non-representational phenomena. Encodings require prior representations in order to be defined. But presumably there were no representations at the moment of the Big Bang origin of the universe, and there are representations now. Representation has emerged out of non-representation, and, since that is impossible for encodings, encodings cannot be the basic form of representation (Bickhard, 1991, 1993).

The assumption, nevertheless, is that there *must* be a solution — after all, representation clearly does exist, and what else could there be but encodings? — and work continues. I am proposing that this programmatic approach to representation is logically impossible, and, therefore, that such work is ultimately fruitless.

Encodings are representations by virtue of *defined* correspondences *with representations*. The general approach assumes that representation can be understood in terms of correspondence (of some sort) with *that which is to be represented*. The problem, as with foundational encodings, is to model how any correspondence could yield representational content of what is to be represented — of what is on the other end of the presumed encoding correspondence.

What kind of correspondence? There is no problem about the *existence* of correspondence. The universe contains myriads of them, of many different kinds: factual, causal, informational, nomological, and so on. All of these have been proposed as candidates for the *special* kind of correspondence that, unlike all the others, yields or constitutes representation at one end of the correspondence about whatever is on the other end of the correspondence. All have failed; some have been abandoned; some are still pursued. To assume that there is some special sort of correspondence with such unique representational power is to assume that there is some special sort of correspondence that carries representational information about its existence and about the other end of the correspondence.

There is, in fact, such a form of correspondence. Morse code is an example. But this form of encoding correspondence, the only genuine form, *presupposes* ALL of the epistemic issues that computationalism wishes to solve with encodings.

In-principle arguments. There is not space to rehearse the numerous attempts that have been made, and that are currently proposed, to find this special form of correspondence — and to recount all of the ways in which they have failed. Programmatic approaches to anything, including to representation, are capable of unbounded variation on the theme, with the continued hope and promise that the next attempt might do it. Only in-principle arguments can show a programmatic approach to be flawed, as in Chomsky's demonstration that ANY model constructed within the strictures of associationism will be necessarily inadequate to the facts of language. I have offered some in-principle arguments against encodingism above — the assumption that encodings can provide their own foundational representational contents is incoherent — and offer many more elsewhere (Bickhard, 1980, 1992, 1993; Bickhard & Terveen, 1995; Campbell & Bickhard, 1986). Even in-principle arguments, however, are not persuasive if there is no known alternative — "What else could there be?" and maybe the in-principle arguments are wrong.

Tangles of red-herrings. Before turning to the conception of representation that I wish to offer as an alternative, however, I will illustrate with one example the sorts of logical and conceptual tangles that can be encountered in attempting to fulfill an encodingist approach to representation. Consider the disjunction problem (Fodor, 1987, 1990; Loewer and Rey, 1991). This problem arises when it is assumed that the special representation-constituting form of correspondence might be one that arises from a causal connection from the “to be represented” to the “representation by correspondence”. Such causal-connection correspondences are a natural possibility to pursue, since one paradigm for the investigation of representation is perception — perception construed as the processing of (and in) causally related chains of inputs from world through sensory transducers to the brain.

How is error possible? One problem that this approach gives rise to, however, is the problem of the possibility of error. In particular, if the causal relationships — the special form of correspondence relationship — *exists*, then the “to be represented” is *in fact* present and has in fact initiated the proper causal process, and the representational correspondence, therefore, is *necessarily* correct. On the other hand, if any part of this initiation, causal transmission, transduction, processing, or whatever else that is taken to be crucial to the representation-constituting special form of correspondence is absent or deficient, then the special correspondence does *not* exist, and, therefore, the representation doesn’t exist. In particular, there is no representation in existence that could be *not correct*. The problem, then, is to explain how representation could even *possibly* be in error, in this view.

The disjunction problem. One version of this problem is the disjunction problem. This problem turns on the question of how it can be determined which correspondences are in fact the right correspondences to be constitutive of correct representation, and how they can be distinguished from those that constitute errorful representations. If a “cow” representation is normally evoked by cows, that is its correct and desired function. But if it is evoked by, say, a horse on some dark night, we would like to be able to say that that is an error. The disjunction problem poses the question: How can we avoid the conclusion that what had been taken as a “cow” representation is really a representation of the disjunction “either a cow *or* a horse on a dark night”?

Note that this is a problem from *within* the basic correspondence framework: *assuming* that representation equals correspondence (of some sort), how can we distinguish the correspondences that constitute correct representations from those that constitute false representations?

Asymmetric dependency. One proposed solution to this problem is “asymmetric dependency” (Fodor, 1987, 1990; Loewer and Rey, 1991). The idea of this “solution” is that there is an asymmetric relationship between the correct and the incorrect cases — between cows and horses on dark nights — and that that asymmetry suffices to distinguish them. The proposed asymmetry is one of counterfactual dependence of the *possibility* of the one kind of correspondence relative to the other. In particular, it is claimed that “horse on dark night” correspondences would never occur unless the “cow” correspondences occurred; conversely, “cow” correspondences could very well occur even if the “horse on a dark night” correspondences never did. The idea is that the correct correspondence — with “cow” in this case — is the critical one for the very existence of the (representation constituting) correspondence relationship in the first place, so any instances of error must be in some sense parasitic on those correct versions.

There are a number of challenges that can be made to this solution. I offer two here. They have the form of counterexamples. First, a transmitter molecule docking on a receptor in a cell surface, and thereby triggering appropriate activities internal to the cell, is an example of a causal chain correspondence (transmitter to internal activity), and, furthermore, a causal chain correspondence that is followed by appropriate functional and evolved functional activity (Bickhard, 1993). Yet, there is at best a functional story to be told here. There is no epistemic relationship between the internal cell activities and the transmitter molecule; the cell doesn't contain a representation of that molecule, nor of the activities of the previous cell that created that molecule, nor the activities that released it, and so on. The asymmetric dependency proposal, in other words, presupposes that *some* correspondence does in fact constitute representation. The transmitter molecule challenges that assumption: how can the genuine representation-constituting correspondence be distinguished from the transmitter correspondence?

One reply might be that representational correspondences are capable of asymmetric dependencies. So, only those that are capable of such asymmetries are representational at all. The asymmetries, in turn, distinguish within those representational correspondences between those that are correct and those that are false. Consider now my second counterexample: a poison molecule mimics that transmitter molecule, docks on the cell surface receptor, and inappropriately triggers those internal cell activities. Here is also a correspondence. Furthermore, it is a type of correspondence that is asymmetrically dependent on that of the transmitter molecule: the transmitter correspondences could exist quite nicely even if the poison correspondences never did, or never could, but the poison correspondences are *only* possible because of the transmitter correspondences. Yet, here again, we have only a functional story. Just as there is no representation in the cell that accepts that transmitter molecule, there is no *false* representation in the cell that accepts the poison molecule. The asymmetric dependency relationship captures at best a functional asymmetry, and it can be modeled at a strictly functional level. It suffices neither to characterize true representation from false representation, nor representational correspondence from non-representational correspondence (Bickhard, 1993).

More tangle. The disjunction problem, and its asymmetric dependency proposal as a solution, is but one of many problems about and within the encodingist framework. A sampling of others: How could we check our encodings for correctness, if we can only check them against themselves — we have no other access to what they are supposed to represent other than those encodings per se? How could we ever construct our encoding copies of the world without already knowing the world that we are attempting to copy? If we cannot encounter error in our checking of encodings, how can we ever control learning processes? Any element in correspondence with something in the world will necessarily also be in correspondence with thousands of other things in the world — not just the chair, but also the light patterns, activities in the retina, electron orbital activities in the surface of the chair, the motion that placed the chair at that position, the construction of the chair, etc. etc. — which of these correspondences is the representational one? As mentioned above, How could representation come into being out of non-representation in the history of the universe? And on and on (Bickhard, 1991, 1992, 1993; Bickhard & Terveen, 1995). None of these are solvable or resolvable from within the encodingist approach.

Part of the reason for rehearsing this problem and purported solution and failure of that purported solution — and mentioning the host of additional problems — is to illustrate how tangled these investigations can be; and to give some sense of how easy it can be to become lost in these tangles — looking for a way around this tiny barrier or looking for how to fill that small hole in the argument. A second reason for this overview of the disjunction problem and asymmetric dependency proposal is to illustrate that many, perhaps most, of the problems that occupy the efforts of work within the encodingist, representation as correspondence, approaches are simply red herrings (Bickhard, 1993). They are problematic ONLY because of the approach taken, and do not offer fundamental problems for the alternative approach that I suggest. In particular, the possibility of "being in error" is trivially present for the model of representation that I urge. I turn to it now.

Interactivism

The alternative model of representation is called interactivism. Interactivism models representation, and representational content, as particular properties of systems that interact with their environments, rather than as particular kinds of correspondences. By beginning with fundamentally different framework, interactivism simply never encounters most of the problems of encodingism.

Differentiation and implicit definition. Consider a system interacting with an environment. In part, the internal processes of the system that participate in that interaction will depend on the functional organization of the system that is engaged in the interaction. In part, however, the internal processes, and course of those internal processes, will depend on the environment being interacted with. Similarly, the internal condition that the system is in when that interaction has ended — when that subsystem ceases its processing, for example — will depend in part on the environment with which the interaction occurred. If there are two or more possible such internal conditions that the system could end up in — two or more possible final states of the interactive subsystem — then those states will serve to differentiate classes of possible environments, and the particular final state of a particular interaction will differentiate the particular environment that the interaction was with. So, if the subsystem ends up in final state "**A**", say, then it has encountered the type of environment that yields final state "**A**" when engaged in interaction, and similarly for final state "**B**". In effect, the possible final states implicitly define the classes of environments that would yield them, and actual interactions classify environments among those implicitly defined classes.

It is important to note that the differentiations involved here do not create nor constitute any representational content about the environments, or classes of environments, other than their property of yielding this final state rather than that final state. In particular, no such interactive differentiator could generate encodings, because there is no representational content about the environments differentiated.

Nevertheless, there will be some *factual* properties of environments that underlie their tendency to yield some final state, and a differentiation will, therefore, generate a *factual* correspondence with whatever those factual properties may be. That is, an interactive differentiator will create *precisely* the sort of correspondences that are classically taken to be constitutive of encoding representations. This is even more obvious in the case of an interactive differentiator that is simplified to the point of having no outputs — a differentiation based entirely on the processing of inputs, such as in sensory "transduction" (or connectionist "pattern recognition"). The same point holds: such processing can be sufficient to differentiate categories of possible environments (or input patterns) — those that yield this internal outcome rather than that internal outcome — but they do not and cannot yield any representation about those (categories of) environments. The differentiations are open as to what is in fact being differentiated; the definitions of the categories are implicit, not explicit.

Interactive differentiators, then, model what is standardly construed as encodings. But the interactive model makes no claims that these correspondences somehow magically constitute representations of what has been differentiated. That, of course, raises the next question: how *does* the interactive model account for representational content?

Representational content. Consider an interactive differentiator subsystem in the context of its broader system. It might be useful for that broader system to make use of the final states generated by the subsystem in determining the future course of overall system interactive activity. If, for example, the subsystem generated final state "**A**", that might indicate that, among other things, subsystem "**S2**" would, if engaged in interaction, yield final state "**Q**", and that subsystem "**S22**" would yield final state "**R**". Such final states can be constituted as functional or biological conditions in the system that could be sought in a goal directed sense by that system: so, the possible final states that could indicated as possibilities by an initial final state "**A**" — upon appropriate intermediate further interaction, such as "**S22**" — potentially play an important role in the system satisfying its goals. Final state "**A**", for example, might be the outcome of a visual scan that internally indicates, among other possibilities, that, if a certain further interaction were engaged in, "**S22**", a further final state, "**R**", of raised blood sugar would be reached. If raising blood sugar is an active homeostatic goal, then that possibility might be selected as a next activity of the system.

The general point is that internal system indications of what further interactions would be *possible* for the system to engage in, and what further possibilities would follow if they *were* engaged in, can be useful to the system. We should expect, in fact, potentially quite complex organizations of such indications of potentiality which would yield further potentialities, which would yield still further potentialities, and so on, in complex interactive systems — such as living systems.

The model to this point is strictly functional. No claims have yet been made concerning representation or representational content. In fact, a typical move of claiming representation for input processing correspondences has been explicitly eschewed. Nevertheless, the model does now contain a functional property that emergently constitutes representation. This is representation in a most minimal sense, but representation nevertheless.

The minimal property of representation is that of the possibility of being *false* from the perspective of the system itself, and the possibility of being *falsified* by the system itself. More generally, the possibility of having some truth value for the system itself. The caveat concerning "the system itself" derives from the common practice of attempting such analyses from the perspective of some human observer or analyzer of the system, and smuggling the observer's perspectives and interpretations into the purported model of the system. If a human scientist, for example, were to be using some internal cell activity as a signal, an "encoding", of some prior activity which yielded the manufacture and release of a transmitter, then the poison molecule would generate a false representation *for the scientist*. At issue, however, is the sorts of representations in the *scientist-as-system* per se, not the representational usages that the scientist can derivatively make of aspects of other systems. That is, at issue is primary representationality, not derivative representationality.

The indication of the potentiality of further interaction in a system is an indication that has truth conditions *external to the system*. The indicated potentiality may in fact not be a potentiality, and it will be a potentiality only if the world in general supports the indicative relationships involved. Only if the world is in fact a place in which, whenever "this interaction yields this outcome", then "that interaction would yield that outcome" will an indication of that relationship hold. So, the indication, which is strictly functional and strictly internal to the system, depends on the world for its truth conditions. Note that such an indication *has* external truth conditions, but it does not *represent* those truth conditions.

Furthermore, if such an indication is false, there is a possibility for the system itself to discover that. If the system does in fact engage in an interaction, reaches a particular outcome, then engages in an indicated potential *further* interaction, and that further interaction fails to complete, or completes with a non-indicated final state, then the initial indication is false *for the system itself*. Indications of interactive potentiality are, on the one hand, functional indications, but, on the other hand, they have truth conditions testable by the system itself. Such indications are the primitive form of the emergence of interactive representation.

Indications of potential interactions constitute an emergence of minimal representationality, in the sense of having truth values determinable by the system itself. They do not, however, look much like representations of our familiar world of objects in space and time, related causally, with other agents, language, and so on. They are a form of representation most readily found in primitive systems, such as paramecia.

From action to objects. Nevertheless, the claim is that this is the primitive form of representation out of which all others are constructed. That claim is a programmatic claim that can only be demonstrated progressively via presenting the models of such constructions (see below). Models of the nature of learning, emotions, consciousness, language, perception, memory, values, rationality, development, and others have, in fact, been offered (Bickhard, 1980, 1991b, 1992, 1992b, 1992c, 1993; Bickhard & Campbell, 1992; Bickhard & Richie, 1983; Campbell & Bickhard, 1986, 1992). Obviously, much more work remains. The critical point for my current purposes, however, is that interactivism *does* constitute an alternative to the standard correspondence notions of representation — the notions of representation upon which computationalism is founded.

Pragmatism. Interactivism is a model of the emergence of representation out of action and interaction, rather than as a form of correspondence. As such, it participates in a tradition that has looked to action for representation — pragmatism. There are many differences of detail and some fundamental differences between interactivism and classical pragmatisms. The basic intuition, however, that representation is not a static correspondence, but instead is a functional emergent of action is in common.

Piaget and objects. Furthermore, interactivism connects strongly with generally Piagetian ideas, from within the pragmatist tradition (Peirce, James, Baldwin, Piaget), of how our world of objects, and so on, *can* be constructed out of action (Piaget, 1954). Roughly, for example, objects, from a child epistemological perspective, are complex patternings of interactive potentialities — visual scans, manipulations, translations, throwings, chewings, etc. — that remain *invariant as overall patterns* under large numbers of possible interactions. None of visual scans, manipulations, etc., for example, change the overall pattern of interactive potentialities that *is* that object for the child. Burning or crushing, on the other hand, would change those potentialities: manipulable objects are generally not invariant in their interactive potentialities under burning or crushing.

Emergent representation; false representation. Interactive representation is emergent in relatively simply interactive system organization. Most importantly, it is emergent in organizations of system organization that need not themselves be representational at all. Interactive representation is emergent *de novo* — it does not require that we already have representation in order to get representation, as does encodingism. Consequently, interactivism does not encounter the incoherence of encodingism in accounting for the representational content of its grounding elements — the interactive ground is not one of elements at all, and the representational content *is* the indicated potentiality of interaction. It does not have to be provided from anywhere else (Bickhard, 1991, 1993).

Furthermore, interactive indications are trivially capable of being false, and of being falsified. There is no problematic of the disjunction problem, and no necessity for gyrations such as the asymmetric dependencies. These are red herrings that simply do not arise from within the interactive framework. These are symptoms of the general approach to representation in terms of correspondence, not fundamental problems about representation per se (Bickhard, 1991; Bickhard & Terveen, 1995).

The Dominance of Computationalism

Computationalism is an approach to cognitive science that is deeply embedded in the encodingist perspective. Newell's symbol system hypothesis, for example, is a straightforward, even somewhat naive, statement of encodingism, yet it is proposed and accepted as the core foundation of cognitive science and artificial intelligence (Newell, 1980; Bickhard & Terveen, 1995), including of the major project of SOAR, for which has been claimed such wondrous properties as reflection, learning, intelligence, and other mental phenomena (Laird, Newell, Rosenbloom, 1987; Laird, Rosenbloom, Newell, 1986; Rosenbloom, Laird, Newell, 1988; Rosenbloom, Laird, Newell, McCarl, 1991; Bickhard & Terveen, 1995). Lenat's CYC project, attempting to construct a truly gigantic knowledge base of encodings — millions and millions of "facts" requiring tens of years and millions upon millions of dollars — on the assumption that only problem with previous encodingist efforts is that they have been on too small a scale, is another testament to the power and ubiquity of the encodingist assumption (Lenat, Guha, 1988; Lenat, Guha, Wallace, 1988; Lenat, Feigenbaum, 1991; Bickhard & Terveen, 1995). Encoded propositions, encoded features, encoded entities, encoded rules, completely dominate the efforts to understand the cognitive properties of the mind.

Connectionism. Connectionism offers, in some respects, a major alternative to standard symbol crunching approaches to cognitive science. It offers, among other things, the possibility of training a system to generate internal categorizations of input patterns, instead of having to design a physical transducer or write a program. This has been construed as "learning" and has been part of the source of the excitement associated with connectionist approaches. There is much to analyze about connectionist approaches, both strengths and weaknesses, but, with respect to the basic issue of encodingism the issue is clear: connectionism offers no alternative to standard encodingist assumptions concerning the nature of representation. The distributed patterns of activation that are taken as "representations" of input patterns in connectionist systems are simply passive differentiators, in the interactive sense, or non-inferential transducers, in the symbol manipulation sense, that are supposed to represent by virtue of the *correspondences* they have with those input patterns. Connectionist "representations" are purported encodings (Bickhard & Terveen, 1995).

Computationalism Requires Encodingism. If encodingism fails, then computationalism necessarily fails. Without encodingism, computationalism has no claim to be able to address the fundamental problem of representation. Absent representation, computationalism has no claim to be able to address *any* intentional properties.

Summary

I have argued that encodingism is incoherent, and that computationalism does therefore fail. It is in serious trouble indeed. But a naturalism of mind — in which mind is addressed as constituted of natural phenomena in the world, rather than as some supernatural intrusion into the world — is not precluded by these arguments. Interactivism, in fact, offers an alternative naturalistic approach to understanding representational phenomena. Interactivism offers a naturalism of representation that is not subject to root incoherences, impossibilities of evolutionary and developmental emergence, impossibilities of error, and so on (Bickhard, 1992, 1992c, 1993). Interactivism offers a *natural* approach to the naturalistic emergence of representation out of non-representation.

REFERENCES

Anderson, J. R. (1983). *The Architecture of Cognition*. Cambridge: Harvard University Press.

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

Bickhard, M. H. (1991). The Import of Fodor's Anticonstructivist Arguments. In L. Steffe (Ed.) *Epistemological Foundations of Mathematical Experience*. New York: Springer-Verlag, 14-25.

Bickhard, M. H. (1991b). A Pre-Logical Model of Rationality. In L. Steffe (Ed.) *Epistemological Foundations of Mathematical Experience* New York: Springer-Verlag, 68-77.

Bickhard, M. H. (1992). How Does the Environment Affect the Person? Invited chapter in L. T. Winegar, J. Valsiner (Eds.) *Children's Development within Social Contexts: Metatheory and Theory.* Erlbaum, 63-92.

Bickhard, M. H. (1992b). Commentary on the Age 4 Transition. *Human Development*, 182-192.

Bickhard, M. H. (1992c). Levels of Representationality. *Conference on The Science of Cognition*. Santa Fe, New Mexico, June 15-18.

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

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., Richie, D. M. (1983). *On the Nature of Representation: A Case Study of James J. Gibson's Theory of Perception*. New York: Praeger.

Bickhard, M. H., Terveen, L. (1995). *Foundational Issues in Artificial Intelligence and Cognitive Science: Impasse and Solution*. Elsevier.

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

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. (1992). Types of Constraints on Development: An Interactivist Approach. *Developmental Review, 12*(3), 311-338.

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

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

Fodor, J. A. (1990). *A Theory of Content and Other Essays*. MIT.

Harnad, S. (1987). Category Induction and Representation. In S. Harnad (Ed.) *Categorical Perception*. pp. 535-565. Cambridge: Cambridge University Press.

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

Laird, J. E., Rosenbloom, P. S., Newell, A. (1986). Chunking in SOAR: The Anatomy of a General Learning Mechanism. *Machine Learning, 1*(1), Jan 86, 11-46.

Lenat, D. B., Feigenbaum, E. A. (1991). On the Thresholds of Knowledge. *Artificial Intelligence, 47*(1-3), 185-250.

Lenat, D., Guha, R. (1988). The World According to CYC. MCC Technical Report No. ACA-AI-300-88.

Lenat, D., Guha, R., Wallace, D. (1988). The CycL Representation Language. MCC Technical Report No. ACA-AI-302-88.

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

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

Palmer, S. E. (1978). Fundamental aspects of cognitive representation. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization*. Hillsdale, NJ: Erlbaum.

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

Rosenbloom, P. S., Laird, J. E., Newell, A., McCarl, R. (1991). A Preliminary Analysis of the SOAR Architecture as a Basis for General Intelligence. *Artificial Intelligence, 47*(1-3), 289-325.

Rosenbloom, P., Laird, J., Newell, A. (1988). Meta-Levels in SOAR. In P. Maes, D. Nardi (Eds.) *Meta-Level Architectures and Reflection.* Elsevier, 227-240.

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