Are we wrong about representation?
An extended essay review of Bickhard & Terveen (1995)

To appear in the *Journal of Experimental and Theoretical Artificial Intelligence*

CLAYTON T. MORRISON
Department of Philosophy (PACCS)
Hinman 129
Binghamton University
Binghamton, NY 13902-6000
Phone (PACCS Program): (607) 777-2446
e-mail: `clayton@turing.paccs.binghamton.edu`

Book Review:

Foundational Issues in Artificial Intelligence and Cognitive Science: Impasse and Solution

by Mark H. Bickhard and Loren Terveen

MARK H. BICKHARD and LOREN TERVEEN. (1995). *Foundational Issues in Artificial Intelligence and Cognitive Science - Impasse and Solution*. Elsevier Scientific. xii + 384 pp. (pbk) US $97.00, ISBN 0-444-82520-7, (hbk) US $150.00, ISBN 0-444-82048-5

[In the US/Canada orders may be placed with: Elsevier Science, P.O. Box 945, New York, NY 10159-0945; Phone (212) 633-3750, Fax (212) 633-3764, E-mail: usorders-f@elsevier.com]

1 - Introduction

In his recent book, *The Elm and The Expert* (1994), Jerry Fodor opens with a rather puzzling remark: “It may be that I have gotten myself into a philosophical situation about which all that can helpfully be said is that I ought not to have gotten myself into it” (p.1). Fodor is referring to what is a paradoxical and deep problem in his theory concerning the relations between representational content, computation and explanation of the mind — it seems that the three may not be compatible at all, at least in the way Fodor wants them to. Fodor, of course, does not plan to give up that easily, and the rest of his book is devoted to trying to get himself out of this dilemma. Mark H. Bickhard and Loren Terveen, however, have quite a different take on the matter. In their recent book, *Foundational Issues in Artificial Intelligence and Cognitive Science* (1995), they respond with an emphatic *yes*: Fodor is indeed in a philosophical “pickle” that he shouldn’t have gotten himself into. The bad news is that Fodor is not alone; in fact, if Bickhard and Terveen are correct, all of cognitive science and artificial intelligence (AI) is at a foundational impasse which stems from a deep problem, of which Fodor’s worry is just one symptom. This problem concerns the fundamental nature of representation.

The capacity to represent is a necessary condition for intelligence — without representation there is no cognition, only accidental success of activity in the world; our deepest scientific understanding of what separates intelligent systems from non-intelligent things hinges on our conception of representation. Bickhard and Terveen seek to do nothing less than expose a fundamental incoherency which they argue has been assumed by seemingly every researcher in cognitive science and AI, despite the many surface dissimilarities of the approaches within these fields. If Bickhard and Terveen are correct, and I believe they are, then this fundamental incoherency threatens the goal of cognitive science and AI to build real intelligent machines and understand naturally occurring intelligent systems.

Nonetheless, Bickhard and Terveen are still very much believers in the cognitive science and AI cause, and their book is not intended to only forecast doom for our cognitive sciences. In addition to their foundational critique, Bickhard and Terveen propose an alternative view of representation which avoids the fundamental incoherency they have exposed, thus suggesting a new framework for the study of cognition and phenomena involving representation. This is not to say that such a solution is achieved without a price: to accept Bickhard and Terveen’s prescriptions, we may have to abandon some of our deepest intuitions about what we think representation is, and also recognize that some of cognitive science and AI’s most accoladed projects are not the successes we thought. Regardless of whether the reader ultimately accepts or rejects Bickhard and Terveen’s framework, this book is certain to generate new ideas and heated discussion — it is well worth the read.

The following is an extended book review of Bickhard and Terveen's important book. My primary focus is to present an outline of the central argument of the book. I will begin with some general comments about the book, followed by the outline of the core argument presented in the book. Along the way I will present a justification for why the book deserves the attention of all cognitive science and AI researchers.

1.1 - General comments about the book

The central task of Bickhard and Terveen's book is to identify a particular misguided, deep assumption concerning the nature of representation — *encodingism* — and develop an alternative model which avoids that assumption — *interactivism*. The book constitutes a very detailed exploration into the intellectual and applied research traditions that currently exist in AI and cognitive science; this exploration succeeds in providing a mapping of where and how these traditions fall with respect to these two fundamental and diametrically opposed approaches to the nature of representation.

Bickhard has been wrestling with these issues of representation in print for about 20 years. While several books and numerous articles are already in print which deal with aspects and entailments of his model (particularly Bickhard 1980, Bickhard & Richie 1983, and Campbell & Bickhard 1986), this book represents the most detailed focus on encodingism, including its identification and entailments for cognitive science and AI. For this reason, this book is the best place to start in becoming acquainted with the encodingism critique and the interactive model.

The book presents a collection arguments which make the case for moving from encodingism to interactivism. Part 1 is devoted to outlining the critique of encodingism. This includes presenting encodingism in its bare form, development of the family of anti-encodingism arguments which demonstrate how encodingism is flawed, and then a taxonomy of different kinds of approaches to representation which assume aspects of encodingism. Part 2 then introduces the alternative interactivist framework, with particular attention paid to how interactivism avoids the problems of encodingism, while still adequately capturing the basic principles of representation. Part 3, the largest section of the book (more than two thirds of its content), consists of an extensive overview and discussion of cognitive science and AI projects and practitioners. A wide range of intellectual ground is covered with the intent of exposing the existence of encodingism and its impact, while also demonstrating where ideas diverge from and converge with interactivism. Two of the chapters in Part 3 are devoted to the discussion of two major facets of cognition: language and learning. The chapters include discussion of the impact of encodingism on the modelling these phenomena, and how these phenomena are alternately conceived of in an interactive framework. Finally, the book concludes in Parts 4 and 5 with a brief discussion of the kinds of system architectures that can fruitfully model interactive systems.

The book's examination of the large and diverse fields of cognitive science and AI has two important properties. The first is that Bickhard and Terveen have done an admirably thorough job of tackling such a vast project: a fundamental critique of how cognitive science and AI approach the nature of representation. We as practitioners of cognitive science and AI expect good arguments to back up strong claims, such as: *all* of cognitive science and AI is currently foundering at an impasse.[1] This is where Bickhard and Terveen's survey of the field pays off. Just about any practitioner of cognitive science and AI should be able to find some section which considers topics or people which are relevant to their own work (this is also evidence of how the nature of representation is fundamental to so many facets of cognitive science and AI studies). It is impossible to survey everything, but Bickhard and Terveen have been careful to choose those which are arguably the most representative positions and projects.

Second, the examination also allows Bickhard and Terveen to explore the numerous subtle but important entailments of encodingism (e.g., explaining the difference between "representation as function" and "the function of representation" - p.193). There are many different versions of encodingism and their relations are sometimes non-obvious. Also, not all projects and programs assume encodingist positions for the same reasons. The treatment of each project and position is done with care in order to expose how it may assume aspects of encodingism, or fails to provide crucial accounts which would be required in order to avoid encodingism. Along the way, Bickhard and Terveen also develop key notions for the alternative interactive framework — many of Bickhard and Terveen's results can not be satisfactorily explained until seen in contrast to current approaches, so the comparisons are very effective.

One of the great strengths of the book is that in its wide-spread coverage of the field, with its diagnosis of encodingism and tracing of the implications of the encodingism critique, a much better picture emerges of the problems with encodingism and a feel for what exactly is wrong. This trek through the literature also gives new perspectives on many issues, including new perspective on several already observed but seemingly unrelated "problems." For example, Bickhard and Terveen agree with Searle's (1980) basic critique in the Chinese Room argument, but not with his conclusions. By and large, Bickhard and Terveen agree with the "robot reply" to Searle's problem, but make the case that the most important feature missing in Searle's account is *timing* for skilled interaction as part of representation

[1] One obvious question to ask is: Why are Bickhard and Terveen the only ones who can see this "impasse"? The answer, it turns out, is that they aren't. In their book they survey the work of a number of researchers who have raised similar issues and concerns with standard approaches to representations, and their respective critiques are compared and contrasted with encodingism and interactivism.

— a point which until now had not been highlighted. The following is a sampling of illustrative projects, people, and topics addressed in the book:

Agre & Chapman	Frame Problem(s)	Jean Piaget
Rodney Brooks	Genetic AI	Procedural Semantics
Central Nervous System Functioning	James J. Gibson	John Searle
Bill Clancey	Steven Harnad	Benny Shanon
Adrian Cussins	Ben Kuipers	Situated Automata Theory
CYC	Language Learning	Brian C. Smith
Hubert Dreyfus	Maturana & Varela	SOAR
Dynamic Systems	Motivation	Tarskian Semantics
Ken Ford	PDP (Connectionism)	Turing Machine Theory
	Perception	

1.2 - Level of difficulty

This book would not be appropriate as an introduction to cognitive science and AI text, unless it was accompanied by a fair amount of supporting text. Full digestion of the book requires a good background in cognitive science and AI, as well as some philosophy. I would certainly recommend it, however, for graduate-level work in cognitive science and AI, particularly because of its survey nature with respect to the issues involved — it is guaranteed to generate much fruitful discussion.

Because the book covers such a large amount intellectual ground, particularly with the goal of pointing out foundational errors in intellectual traditions, the reader should be expected to have some understanding of the tradition being critiqued in order to see the full force of the argument — it would be impossible for Bickhard and Terveen to have done a complete exegesis for all the topics covered. For example, I have only a little familiarity with hermeneutics, so I was not certain as to the correctness of Bickhard and Terveen's characterization of that field and their claim that it has partially recognized issues stemming from encodingism. I found that in general, however, their characterizations were accurate and fair. The writing is straightforward and the arguments are clearly presented so that the most difficult comprehension comes with the inherently complex conceptual issues. The only aspect of the book that is disappointing is its price. Elsevier should have made this book affordable to individual researchers and graduate students, not just academic institutions.

1.3 - Who should read this book?

I recommend this book to any researcher currently involved in cognitive science and AI — particularly those who base their explanations of cognition on the representational power of systems. Also, anyone should read it who wants a good overview of the state of research in semantics and representation in cognitive science and AI. In addition to the large amount of intellectual ground covered, the

book also presents a novel position on representation which differs dramatically from the variety of current approaches to the nature of representation. It also finds a common theme within, and subsequently unifies, a number of previously unrelated programmatic approaches (in particular, the pragmatist school, Gibson's theory of affordances, and Piaget's constructivism). This book is part of a mature body of research that deserves a lot of attention.

2 - Issues in Representation

Before I turn to the central argument in Bickhard and Terveen's book, I will first orient the reader by laying out a quick map of the field of issues regarding representation. The notion of representation funds the hope that cognitive science and AI will someday achieve their ultimate programmatic goals: understanding the mental and building systems which have genuine mentality. The argument for the importance of representation for achieving these goals is simple: for a system to behave intelligently in the world, it must somehow be able to represent its world — that is, some aspect internal to the system must somehow be *about* the world, and the system must somehow be able to take advantage of this "aboutness" to reason and act intelligently in the world. Denying that intelligent systems have this connection with the world would mean that the system could not make good decisions about what to do based on properties of the world that were not proximately causal because the system would have no access to (and no representations of) those properties. What, then, would be the distinction between humans (and other "intelligent" animals) and other naturally occurring phenomena (like rocks, rivers, or the weather)? Representation just *is* the difference between these intelligent and non-intelligent phenomena. If you accept this argument for representations, then you're playing the game of cognitive science: you accept that representations exist and that they are important for explaining intelligence.

Much of the discussion of representation has centered around the *form* of representation. Two general issues concerning the form of representation can be characterized (following Cummins 1996, p.1) as what kind or kinds of representational schemes intelligent systems employ (e.g., images, symbolic structures, activation vectors, dynamic system trajectories), and which of the available forms within a given scheme are employed in a given task. Over the past decade, discussion in the literature concerning the forms of representation has focussed on the debate between "traditional" symbol manipulation and parallel distributed processing (PDP or connectionism; see Clark (1993), Fodor & Pylyshyn (1988), Smolensky (1988), and van Gelder (1991)).

Another issue of representation is that of representational *content*. This concerns the questions of "in virtue of what does something represent?" and "what is the nature of this representational property — the *aboutness* of representation?" Most of the attention given to this issue has been philosophical, and a variety of approaches have been espoused, including straight-forward correspondence theories, theories about relations between internal system functions, attempts at grounding content causally, and attempts at grounding content in terms of evolutionary function (see Block (1986), Cummins (1996), Dennett (1987), Dretske (1981, 1986), Fodor (1994), Harnad (1990), Haugeland (1985), Lloyd (1989), Millikan (1984, 1993), and Newell (1980)).

There has even been voiced the desire to do away with representation talk altogether, under the general argument that our current theories of representation have been misleading and we would do much better not hindering ourselves with their pitfalls and blind alleys (see Brooks (1991), Thelen & Smith (1994), and van Gelder & Port (1996)).

Stances towards both the form and content of representation (or no representation at all) will entail certain fundamental positions on what the ultimate nature of representation is (or is not). At the same time, there are also "levels" of dependency among theoretical assumptions concerning representation, such that two views of representation may disagree about some aspects of the nature of representation, but both still agree in accepting a more fundamental assumption. But what if our most fundamental assumption concerning the nature of representation is wrong? What if all current models of representation explicitly or implicitly accepted a basic assumption which is false? (Note that to give up on representation as it is currently conceived and not offer an alternative is dangerously close to assuming that an alternative does not exist.) Such a foundational flaw would have severe entailments for the general outlook of the field.

3 - Basic features of representation

There are two general aspects of representations — these aspects seem to satisfy at least the very minimal set of features that would need to be addressed by any model of representation in order for it to claim to be about "representation" (i.e., these are points about which a theory of representation should say something):

(1) A representation is *about* something. That is, it holds some relationship to something other than itself; and

(2) A representation is *for the system*. This feature of representation is more contentious than the first. It seems that there are two versions: a weak version and a strong version.

The weak version of "representation for the system" posits that whatever a representation in a system is, it somehow participates in the further functioning of that system. In this sense, a representation plays some functional role in the operation of the system. The weak version is intuitively appealing because it suggests that we can come up with criteria by which we can naturally separate those systems which represent (most animals and some machines) from those which don't represent (at least in any interesting way; e.g., rocks, rivers and weather). So, for example, the operating system of a computer network keeps track of the users

on the system and seems to "represent" them because if we remove one of the users, the operating system will respond appropriately to its absence by reallocating memory, updating files, changing its users status, etc. — there is a sense of proper functioning of the system which depends on conditions which the system can detect (represent). The trouble with this weak version is that it is difficult to sort out trivial or incidental causal connections which affect the system from those that actually seem to be doing some representational work (e.g., why isn't a "bug" in the system represented because it causes the system to respond "appropriately" — namely, crash because something is wrong which it can't fix), and there is a lot of work to be done to explain what is meant by "proper function."

The strong version adds a bit more: "representation for the system" is taken to mean that representation is to be characterized in terms of "what *of* the representation's 'aboutness' the system can be said to have 'access' to." Intuitively, think of it in terms of what the representation *is* from "the system's perspective." The strong version is intuitively appealing because it captures the intent of the weak version to separate cases of representing the world from mere causal connection, and it should provide criteria for avoiding the problem of which causal connections are relevant (namely, only those that the system has "access" to or can make use of). Furthermore, it seems somehow to be more "cognitive." In this case, the above example of the operating system may not be strong enough to capture what representation is. The strong version prompts us to ask questions like: "In what sense does the operating system really *know* about 'users' and their presence or absence?" The trouble here is that what it means *for the system* to have *access* to a representation's aboutness (its content) is difficult to make clear without begging interesting questions.

Representation being "about something," and the weak and strong versions of "representation for the system," are by no means exhaustive of the distinctions which may be made. And, a good theory of representation might show how these pose a false dichotomy or simply split up the features of representation inappropriately. However, they are at least intuition pumps and a good place to start so that we have some initial guiding principles for taxonomizing the variety of approaches to representation, and they also outline the kinds of things that a theory of representation should say something about.

4 - A succinct statement of the central argument of the book

Now I am ready to consider Bickhard and Terveen's central argument in their book. Bickhard and Terveen have identified a framework of representation which, while capturing many of the intuitive notions of what it is to represent, is ultimately based on a fundamental incoherency. This framework is called *encodingism*. The fundamental incoherency of encodingism is not just a matter of "conceptual messiness," but would, if assumed, actually threaten the hopes that any research program has of eventually explaining and building intelligent systems — and this includes even barring access to some of the more "practical" goals of research (e.g., making things work). They propose in the place of encodingism an alternate framework, called *interactivism*, which avoids the fundamental problems of encodingism and thus promises to transcend the problems caused by encodingism.

Through an extensive review of the literature in cognitive science and AI, Bickhard and Terveen make the case that most of the field explicitly or implicitly assumes key features of encodingism, and thus draws on the myriad problems and ultimate foundational incoherency which encodingism entails. They conclude that because cognitive science and AI largely assume encodingism, these fields are at a foundational impasse. Their review includes arguments for evidence of this impasse. Their literature review also uncovers that many of the contemporary arguments which are *critical* of current cognitive science and AI approaches constitute partial recognitions of the problems with encodingism and are thus convergent with Bickhard and Terveen's criticism of encodingism. This means that the encodingism critique serves to unify these different criticisms, aiming them at the same fundamental target. Furthermore, Bickhard and Terveen make the case that many of the developments in the field, both in response to these criticisms and based on theoretical and empirical findings, have been heading in the direction of interactivism. Along the way, they develop further entailments of the interactive model and outline a picture of what a new research program following interactivism would entail.

An important note to make here regards how Bickhard and Terveen argue for their identifications of encodingism. The logic of the grand argument of the book is that encodingism is ultimately incoherent because of a basic assumption; interactivism is proposed as an alternative which avoids this incoherency by not accepting that assumption; interactivism also has some other nice features. However, the argument is *not* made that therefore interactivism is *necessarily* the *only* possible alternative. It's important to understand this as often problems of encodingism are identified by Bickhard and Terveen in terms of how an existing research program *diverges* from *interactivism*. Admittedly, it's not clear what another alternative to encodingism besides interactivism would be[2] — but at this point it hasn't been shown that another alternative is not possible. In general, this is not problematic for their point that the programs in question are encodingist (often it is because of a comparison with interactivism, a successful alternative, that latent encodingism becomes more clear), but it should be kept in mind so that the critical reader is not misled into thinking that interactivism is necessarily the only alternative — that argument has not yet been given.

The following is a brief overview of the definition of encodingism and the general problems with it. The core argument starts with the central assumption that encodingism makes, and by a *reductio ad absurdum* argument (showing that the assumption entails clearly unacceptable things), concludes that this assumption must be wrong. Following the argument, I discuss some of the entailments of accepting encodingism and provide an example of several popular approaches which are actually versions of encodingism. I then turn to present the alternative to encodingism: interactivism. All of these presentations are extremely adumbrated and are not substitutes for the fully explored versions presented in Bickhard and Terveen's book. They will hopefully, however, whet the appetite of the reader. Feelings of skepticism are in a way the most important reason to read the book as Bickhard and Terveen have been very thorough in exploring the many variations of possible replies and rejoinders; chances are they have at least attempted to answer questions which might come to mind, but can't be presented here.

[2] In their defense, nevertheless, Bickhard points out that historically only two approaches have been taken at this level: encodingism (which he argues has been with us for a very long time), and pragmatism (which is a century old).

4.1 - The assumption of encodingism

First, two terms need to be defined:

(1) *Representational content*: what it is in the representation that makes it *about* what it represents — what the representation tells the system (this is essentially the same definition given above in Section 3); and

(2) *Epistemic contact*: whatever is supposed to make the functioning of a representation appropriate to the current situation — it is what gives a representation contact with the world, with an *environment*.

The deepest assumption that the encodingist framework makes is that epistemic contact *constitutes* or *provides* representational content. This assumption entails that the basic representational unit is an *encoding*. An encoding is a "stand-in" for whatever it is that it is supposed to represent. That is, it is a "stand-in" precisely because the assumption claims that epistemic contact yields representational content: a representation tells the system what it is a representation of.

An example of a representational scheme which uses this notion of an encoding as a model of representation is that of transduction. Transduction is technically the transformation of forms of energy. The basic idea is that the system transducers (such as sense receptors) receive energy from the environment that is in causal correspondence with things of importance in that environment. The transducers then "transduce" that energy into internal processes or symbols which are then said to represent what externally caused them. In this way, an internal symbol **Y** "stands-in" for an external event **X** if symbol **Y** is created or activated by the transduction of a signal from **X**. Here, the stand-in changes the form and medium of representation, which in turn changes the ways the representation may be manipulated. For example, event **X** might be the presence of light (photon energy), but its transduced stand-in, symbol **Y**, may be electrical or chemical activity in a neural structure, or an atomic unit that could be manipulated (e.g., as a variable value in a computer).

This encoding scheme can be generalized to chains of encoding relationships, so that a symbol **Y** can stand-in for the symbol **X**, and a symbol **Z** can stand-in for **Y** — and this relationship is transitive so that in such a chain, **Z** could stand-in **X** The crucial aspect of encodings is that these 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" (p.14). Thus, in the transduction example, encoded symbol **Y** "borrows" its representational content from the external event **X** — and this "borrowing" comes about from the transduction of the signal from the external event to the activation of the internal symbol.

To reiterate, the encodingist framework assumes that these encodings take epistemic contact to provide or constitute representational content.[3] Thus, an encoding represents a particular thing (carries content of that thing — is about that thing) and somehow informs the system of what that encoding is supposed to represent. In it's naked form, it is *this* assumption which probably draws the greatest attention for being problematic — and it is from here that all the problems stem. The key problems with encodingism turn on this group of related questions regarding content: what is the nature of this content, where does it originally come from, and how is it carried?

To assume a position that is explicitly or implicitly a part of encodingism is to do at least one of the following:

(1) To assume that encodings are the essence of representation

(2) To assume that encodings are a logically independent form of representation (that is, encodings do not rely on any other form of representation in order to be a representation).

Thus, Bickhard and Terveen's project with respect to identifying latent encodingism in the representative research projects they chose involved showing how one or both of these assumptions were made (or indirectly, by showing how a research project takes a characteristic stance towards representation which in turn is shown to make one of these assumptions).

4.2 - The problem with encodingism

Taking an encodingist position, by accepting one or both of points (1) and (2) above, leaves us in a dilemma. Either we: (a) assume that representation is rendered in terms of encodings with representational content, but give no model of *where* this content comes from originally and *how* these elements can carry it; or (b) we attempt to actually explain where representational content comes from and how encodings carry it from within encodingism.

[3] As we shall see, there are other ways of accounting for encodings without the encodingism assumption of stand-ins (with contact constituting or providing content) as the basis for representation; but this, then, won't be encodingism.

Choice (a) lands the programs of cognitive science and AI in a vicious circularity: it is precisely representational content that we want to account for. Cognitive science and AI aspire to explain mental phenomena (cognitive phenomena, the base form of which is representation). Yet, accepting (a) assumes what it is that is to be explained: that encodings have content. Most often, rather than explicitly admitting to choice (a), the question of content is pushed off onto philosophy as a problem which would be nice to have solved, but meanwhile can be ignored so that we can get on to interesting empirical research. The problem is that if such content actually can't be provided within encodingism, then there's good reason to believe that even the practical side of such a research program may be in trouble (certainly conceptually).[4]

Choice (b), while being the more interesting choice because an attempt is actually made to account for content, is ultimately doomed to failure because of encodingism's fundamental incoherency: attempts to explain how encodings carry or originally produce representational content *cannot* be done *within* encodingism. Encodings cannot be the fundamental nature of representation because it is logically impossible to explain by use of encodings alone how it is that they have representational content. Encodings, as construed above, *carry* representational contents, and already established encodings can *provide* representational contents for the formation of some other encoding, but there is no way within encodingism itself for those representational contents to ever arise in the first place. There is no account, and — according to Bickhard and Terveen — no account possible, of the *emergence* of representation.

This point deserves a little more attention. We already know that encodings can transitively pass content along: **Y** can stand-in for **X**, and **Z** can in turn stand-in for **Y**, so that **Z** can then stand-in for **X**. In this chain, **X** is the encoding that is providing content for the other two encodings. At some point, however, this chain of providing content has to stop if we are ever going to capture the *original* provider of content — where it is that representational content comes from. I'll call this the *bottom-level* foundation of logically independent representations. At this bottom-level, representations can't be standing-in for others, and therefore these representations cannot be carrying contents provided by any other representation (to do so would mean that we aren't at the bottom-level, and encodings at this level must be logically independent in order to *be* the bottom-level). So this can't be the

[4] Notice the tension between believing in representations as being *about* something in a real sense, with the presence of such representations making actual natural kind distinctions in nature (e.g., between a child and a bucket of sand), and then claiming that content is a purely "metaphysical" issue. Theoretical proposals of what content is should be thought of as being real theoretical entities, in the same way that possible accounts of what creates the phenomena of heat are; the obvious proposed analogy is that encodings, as encodingism construes them, are like phlogiston.

answer. The only other recourse, however, is to consider this bottom-level encoding as providing its own content. This amounts to asserting "**X** represents whatever it is that **X** represents" or "**X** stands-in for **X**," which doesn't in any way succeed in providing **X** with content — this doesn't make **X** an encoding. This is a vicious circularity. Thus, the assumption that encodings can be original providers of content is incoherent, and we are left with no account of where content comes from or what its nature is. In effect, attempts at choice (b) end up back at choice (a).

The version of this incoherence argument I just presented was from the perspective of attempting to establish a *provider* of representational content from within encodingism. However, the same argument can be made in terms of the necessity of an *interpreter* — in this case, from within encodingism it is impossible to eliminate the interpreter which interprets an encoding as being about something, while at the same time it is the interpreter that we wish to account for. This form of the incoherence argument was adopted from J. J. Gibson's critique of encoding forms of perception (Gibson 1966, 1977, 1979; Bickhard & Richie 1983 discusses this in detail).

4.3 - The result of encodingism's incoherence

The result of the incoherence of encodingism forces an important conclusion. It may help to see this result by temporarily suspending questions of exactly which particular research programs are encodingist (if any), and instead continuing with the following in mind: even *if* you think encodingism is a straw man, but you accept that encodingism is in fact incoherent, you can still agree with the following conclusion — in other words, the issue of which programs are encodingist is, in an important sense, *independent* of whether encodingism itself is incoherent and what that incoherence tells us about representation.

The conclusion is this: since encodingism is incoherent or circular, that entails that the assumptions of encodingism must be wrong: if encodings exist, they cannot be the fundamental nature of representation *and* they cannot be a logically independent form of representation (this is a denial of both assumptions (1) and (2) at the end of Section 4.1). This in turn entails that whatever is the fundamental nature of representation, it cannot be the case that representational content is the same as epistemic contact. Representational content must ultimately emerge in some form other than encodings; once this content has emerged, it could *then* be provided for the constitution of *derivative* encodings.

5 - Considerations and entailments of encodingism

The skeptical reader will probably first consider that encodingism is a straw man. Again, even if that is in fact the case, the above theoretical result still holds (or, at least, it doesn't fall just because it's a straw-man argument). Bickhard and Terveen, however, make a strong case that *all* current cognitive science and AI is locked into varying forms of research programs which ultimately accept the key

assumptions of encodingism, and thus are ultimately led into the above dilemma. This is definitely a very strong claim. Bickhard and Terveen, however, back their claim up by reviewing a large chunk of the representative literature in the field and present arguments along the way for "how much" encodingism is assumed in each. I, of course, will not reproduce all of those arguments here, but I will (in the two sections following this one) make some comments about what constitutes evidence of encodingism and present what I believe are some of the interesting versions of research approaches which, according to Bickhard and Terveen, are encodingist.

To reiterate, the problem of encodingism is not that encodings don't exist at all (except in the sense of being strictly independent of a content provider) — they certainly *do* exist: for example, Morse code, computer languages, traffic signs, written english (note: I mean here the symbols themselves, not necessarily how they are used in discourse), etc. Rather, the problem of encodingism is that encodings cannot be taken to be the *fundamental* nature of representation — encodings that do exist are necessarily dependent on a more fundamental form of representation (for example, the representational capacity of the observer who interprets the encoding).

The fact that derivative encodings exist raises an issue that bears mentioning here. That is, many researchers (including Bickhard and Terveen) believe that there are different kinds of representation, and these kinds are hierarchically arranged in a dependency fashion, the most fundamental form being fundamental to all others.[5] Many of the identifications of encodingism that Bickhard and Terveen make stem from evidence that the model of representation in question is really a form of representation that depends on a more fundamental form, but lacks an account of that fundamental form.

The following question might then be posed: Even if I accept that encodingism is not the fundamental level of representation, why can't I still model certain cognitive phenomena satisfactorily *at* the level of encodings and still be perfectly justified in offering explanations? Which is the appropriate level of representation to use in modelling a particular cognitive phenomena will ultimately depend upon the phenomena to be explained. The rub is that this fundamental level of representation has certain entailments for what any representational phenomena are, and these criteria are inherited by all higher levels. Thus, not accounting for this fundamental level in many cases leaves a model either flatly wrong, or at best an approximation to be replaced by a deeper explanation.

The problem with using encodingism to model many (perhaps most?) levels of cognitive phenomena is that the tasks to which the encodingist-based

[5] It is clear that encodingism likewise assumes such a hierarchical relationship in that the claim is that encodings constitute such a fundamental level: assumption (1) and (2) at the end of Section 4.1 assert that encodings are the "essence" of representation (satisfy the key criteria for being a representation) and are a logically independent form of representation (do not depend on any other form of representation).

representations are put (particularly in terms of explanation, in addition to practical goals) are inappropriate. Attempting to go deeper in explanation within encodingism, however, is impossible (because of the limitations of the incoherency of not being able to account for how representational content arises). Cases where encodings are used inappropriately constitute a kind of category error (this point has also been explicitly made by Clancey 1993). Bickhard and Terveen map out several strands of entailments for particular classes of cognitive phenomena at higher levels of representation which follow from what they believe is the fundamental level of representation. In particular, Bickhard and Terveen devote chapters 11 and 12 to models of language and learning, respectively; I will briefly discuss implications for learning as an example, below.

Raising the issue of modelling cognitive phenomena (the bread and butter of our cognitive science and AI methodology) prompts another very important question to address: Why do encoded projects still seem to work? The answer is that they depend on humans who solve the problems requiring the leap from the non-representational to the "representational," thus doing the hard representational work for them. Encoded projects then, in an important sense, *rely* on the representational power of the humans that created them and use them in order to accomplish their tasks.[6]

For example, Bickhard and Terveen put it nicely in their discussion of SOAR's explicit encodingism: "Laird, Newell, & Rosenbloom are explicit about their user/programmer version of encodingism, stating that SOAR 'encodes its knowledge of the task environment in symbolic structures.' However, to be precise, it is not SOAR that does the actual encoding. Programmers do the actual representational work of encoding a problem in terms of states, goals, operators, and even evaluation metrics" (p.100). In the discussion of "chunking," SOAR's version of learning, Bickhard and Terveen add that: "...just as the programmer must anticipate all potentially relevant objects, features, relationships, atomic actions, etc. to be encoded in the problem space in order to make SOAR function, so also must the programmer anticipate the proper aspects, features, etc. that it might be relevant to ignore or variabilize, and, thus, generalize over. As a form of genuine learning, chunking is extremely weak. From the representational perspective, the programmer does ***all*** the work. To construe this as a 'general learning mechanism' is egregious" (p.104).

Thus, an important question to be asked of systems which claim to be "knowledge bases," or "have" knowledge, is: "For whom is this knowledge?" or "For whom is this a representation?" (Recall the second aspect of representation I gave above in Section 3: representation *for the system*.) For Bickhard & Terveen, the thrust of what is important for explaining cognition is finding an account which answers this question: How does the system itself come to represent or know its world?

A final point that this raises is that encodingism is a *model* of representation. In this sense, it has to do with our *conception* of what we think representation is. Despite what we think, however, if representation is a genuine natural kind, then it already exists, and may already exist in models even if not recognized as such. Bickhard and Terveen are presenting their criticisms in terms of the theoretical models of representation of other research programs. Thus, claiming that a research program is latently or blatantly encodingist does *not* entail that their actual working models do not have genuine representations (which is another explanation for why some models created by an encodingist paradigm work). In fact, Bickhard and Terveen point out that many of these working models do have genuine representations. By in large, these successes, accidental or otherwise, are in robotics — this is because robotics forces the researcher to naturally account for many of the aspects of representation which Bickhard and Terveen argue are necessary. The point to be made is that their creators don't recognize this. And while their creators remain married to an encodingist approach, they will be limited in their scope of theory-driven research and future theory-construction. Also, the successes in producing real representations can't be attributed to the *theoretical* model which "accidently" produced it.

6 - Forms of latent encodingism

I now turn to examine some of the common kinds of research approaches which Bickhard and Terveen argue assume encodingist positions. These are general types of approaches, rather than specific research programs. As Bickhard and Terveen argue, these are attempted solutions to partial recognitions of difficulties within encodingism. There are several different approaches which Bickhard and Terveen have identified, but I have chosen three which I believe are popular, and which would be my first choices for trying to counter Bickhard and Terveen's claims that encodingism is pervasive.

The first is the approach of *observer semantics*. In an attempt to avoid the incoherence of encodingism, this approach moves all basic issues of representation outside of the systems or models being constructed, and leaves the representational issues up to the observer or user of the system to be filled in as required. The observer knows that certain of the inputs, and certain of the outputs, are in such-and-such a correspondence with certain things in the world, and are thus available to be taken by that observer as encodings of those things in the world.

There are two versions of observer semantics that Bickhard and Terveen distinguish: *internal* and *external*. External observer semantics simply considers the relations of the inputs and outputs of the system to things external to the system and disregards any possible representation within the system. Internal

[6] Chalmers, French and Hofstadter (1992) also discuss this issue, claiming that in such cases human researchers have *handcoded* their models so that they behave as representational systems. In many cases, it is this representational work which constitutes the bulk of the cognitive task. Bickhard and Terveen's diagnosis of encodingism explains what this handcoding with respect to representation is at a deep level.

observer semantics, the more "traditional" solution within cognitive science and AI to the problems of encodingism, postulates that there are also correspondences to symbol structures within the system.

Either one of these approaches can have some practical uses in research methodologies. Namely, relying on the observations made from our external observer perspective that a particular part of an organism reacts characteristically to the presence of some stimuli (i.e., *covariation*) is a good indication that that part of the organism is somehow functionally involved in the organism's capacity to respond to its environment. And if the organism in some way represents, then there's a good chance that the characteristic activity in the organism is involved in the processes of that organism's representational capacity. The question is then: Since we are scientists who want to explain how intelligent systems work (and likewise build systems with intelligence), why isn't this enough?"

The main problem with observer semantics is that it is too weak. It does not solve any of the problems of representation, but rather simply avoids them. The representational issues are pushed outside of the system, making phenomena such as the generation of representational content, and intensional stances[7] with regard to representational content, impossible to even address (p.28). Namely, all we could then say from this perspective is, "this system represents because *we*, as observers of the system, can observe the correspondence."[8] Observer semantics pushes the problems onto the observer, but provides no model of how any epistemic observer could possibly "solve" these problems: "If we want to understand observers themselves, we cannot validly do so only by adversion to still further observers" (p.28).

Following the suggestion made above concerning modelling by using encodings, this approach may allow for the *simulation* of certain intensional properties and processes (e.g., those that do in fact involve explicit encoded representational elements in real epistemic systems — although Bickhard and Terveen point out that there is reason to question how commonly this actually occurs). But ultimately, from within this approach, the representational contents are provided from outside the model or system. "Neither the external nor the internal observer semantics view provides *any* approach to the foundational emergence or provision of representational content" (p.29). To stop here would be to give up the goal of a naturalistic account of representation and representational content.

[7] "Intensional" is meant, here, in the sense of "propositional attitudes". An external observer semantics can't get content, and it can't obtain believing, commanding, wishing, fearing, etc. content either.

[8] For the moment, hold any responses about there being a causal connection or factual correspondence; I'll address these moves in the third set of responses, below.

A second possible response, which might follow from taking the above arguments seriously, is to conclude some form of *content innatism*.[9] This results from accepting that within encodingism it is impossible to create an encoding with new representational content. "At best, derivative encodings can be constructed that stand-in for new combinations of already present encodings" (p.25). But as Bickhard and Terveen point out, "...this implies that an epistemic system is intrinsically limited to some basic set of encodings and the possible combinations thereof. That is, the combinatoric space defined by a set of basic encoding generators constitutes the entire possible representational world of an epistemic system. Because that basic generating set of independent encodings cannot be itself generated by any known model of learning, so the reasoning goes, it must be genetically innate; the basic set of encoding representations must have been constructed by evolution..." (p.25).

The central problem with this reasoning is that the problem with encodings is logical in nature (namely, it is based on a logical incoherence) and cannot be solved by evolution any better than it can be solved by individual development, as long as encodingism is assumed true. The argument for innatism assumes that the only way to generate new representations is by combining things that are already representations; but, if this is true, then this would also constrain evolution just as much as it would learning and development. Thus, if evolution *does* have a mechanism by which it could avoid the basic incoherences of encodingism (i.e., if evolution could generate emergent encodings), then it must be via some process that is fundamentally *not* encodingism. In general, this mechanism is not specified.[10] And if it were specified, then an additional argument would be required for why that mechanism is impossible for individual learning and development. Therefore, the assumption that the problem of where encodingist representational content comes from *can* be pushed off onto evolution invalidates the whole argument that supposedly yields innatism in the first place — that is, it contradicts the original assumption of innatism that the only way to generate new representations is by combining things that are already representations. (p.26)

And finally, a third set of possible responses might be attempted — these are perhaps the most common. These responses include trying to ground content in

[9] This might alternately be considered a "strong innatism," or "innatism in-principle," to distinguish it from the weaker version of innatism, which only posits that evolutionary processes do shape phenotypic features (we inherit the potential for features through our genes, so these potentials for inherited features are "innate"). Interactivism, Bickhard and Terveen's alternative to encodingism, does not preclude the possibility of innate content, but entails that arguments (like Fodor's) that *all* content *must* be innate are invalid.

[10] Note: proposing that the mechanism is *variation and selection* is an appropriate response (and Bickhard and Terveen argue that variation and selection is, in an important sense, the fundamental mechanism of learning in the limiting case where there is no prior knowledge); but then, how are organisms born with new contents in the first place so that variation and selection processes can work on them? Within encodingism, even variation and selection processes are constrained to work on some basic set of encodings.

terms of straight factual correspondence, information, covariation, or transduction. These all have the advantage over simple observer semantics in the sense that they actually attempt to explicitly address how content arises without the necessity of an observer to supply it; however, they are also found to have deep problems as well. They are all part of the same general problem in the sense that they all confuse factual correspondence with epistemic correspondence (a correspondence that the system itself knows about). Just because some aspect of a system is in factual correspondence with some property of the environment does not entail that that factual correspondence is somehow available to representation so that the representation makes some epistemic contact with the other end of that correspondence.

I have already introduced how the transduction model works (back in Section 4.1, as the example of an encoding approach to representation), so I will consider it here as representative of this third kind of approach (similar arguments are made by Bickhard and Terveen against the other variations). To review, the transduction model posits that system transducers receive energy from the environment that is in causal correspondence with things of importance in that environment. The transducers then "transduce" that energy into internal encodings of those things of importance in the environment. The claim is then that higher order and more important derivative encodings can then be generated out of this foundation of low-level encodings.

Bickhard and Terveen argue that, "what is overlooked in such an approach is that the only thing an energy transduction produces is a *causal* correspondence with impinging energy — it does not produce any *epistemic* correspondence at all" (p.31). Although transduction may produce correspondences, it does not produce any knowledge on the part of the system of either the existence of such correspondences or of what the correspondences are with — that is, it produces *no content*. This point generalizes in that any correspondence taken as being epistemic (such that in virtue of the correspondence the system is said to "know" something) begs the following question: How does the system know what's on the other end of that correspondence? Within encodingism, the only recourse is to further correspondences.

> "Information, covariation, transduction and correspondence are all *not* content. An element **X** being in some sort of informational or covariational or transduction or correspondence relationship with **Q** might be *one* condition under which it would be useful to a system for **X** to carry representational content of or about **Q**, but those relationships do not provide that content. Content has to be of some different nature, and come from somewhere else" (p.33).

To reiterate what I stated at the beginning of this section, all three of the above kinds of attempts to avoid encodingism's incoherences from within encodingism are *not* arguments that *particular* research programs in cognitive

science and AI are encodingist — I will leave such arguments up to Bickhard and Terveen; and they provide a number of them, treating each variation and project in turn. However, I do think it is interesting to note that these schemes are natural entailments of attempts to answer encodingism and they do naturally converge with a number of methodological and theoretical stances taken in actual cognitive science and AI projects (the Physical Symbol System Hypothesis and several versions of Fodor's semantics are examples).

7 - An example of the implications of encodingism: learning

The general consequences of the encodingism impasse, according to Bickhard and Terveen, are that perception, cognition and language, the "backbone" of general cognitive phenomena as classically understood, cannot be fully understood. On the other hand, many research goals, especially many practical ones within AI, do not necessarily depend on that programmatic assumption of encodingism. Even if it does turn out that core cognitive science and AI is encodingist, that does not entail that the main explanatory goals are likewise incoherent or misguided. The question I will address here, then, is: What damage does encodingism do to our ability to account for certain cognitive phenomena?

An aspect of cognition that would certainly be hindered, if not made impossible, by encodingism is that of learning. Bickhard and Terveen outline three deep problems that encodingism has which jeopardize "genuine" learning: (1) the impossibility of the construction of fundamentally new representations; (2) error criteria and error signals must be predefined for the system; and (3) even with such pre-definitions, the system cannot be just an information processor — it must generate interactive outputs in order to generate feedback. (These last two points constitute missing an account of error *for the system*.) In their account of what encodingism lacks emerges Bickhard and Terveen's own assumptions about what "genuine" learning is — and although these are likewise assumptions, they seem to be ones that we'd like to keep. These also form the foundation for several key properties of the interactive model, outlined in the next section.

The first assumption made is that genuine learning involves the construction of fundamentally new representations. This seems plausible and desirable; if this were not the case, then we would be stuck in some form of the above outlined innatism. Encodingist approaches are stuck in such a position because they have no way of possibly accounting for the emergence of entirely novel representational primitives.

Amongst several reasons for why this kind of innatism is not desired *a priori* is the fact that it assumes that all possible representations-to-be-learned must be anticipated in the combinatoric space of the generating set of basic encodings. This anticipation, in turn, must be done by the designer in the case of AI learning research, and by evolution in the case of human beings and other naturally occurring representors. This leads to strange claims, such as our capacity to represent the contents of concepts like sub-atomic particles, the band *Pearl Jam*, and the particular humor found in *Monte Python's Quest for the Holy Grail* having been selected for (directly or indirectly) by some sort of evolutionary process. While this does not serve as a convincing reductio for everyone[11], it seems conceptually neater to find a way to not require this at all. And, as shown above, *within* encodingism there is still no account of how even evolution could *produce* such primitives. Projects which attempt to find such all-encompassing primitives don't seem likely to succeed (as Bickhard and Terveen argue in their critique of the *CYC* project, pp.107-118).

Another entailment of encodingism's restriction to a combinatoric space of encodings is that learning within the encoding approach, in effect, can be no more than a different principle by which the combinatoric space is explored. Thus, the very distinction between learning and other "search" processes boils down to a difference in the combinatorial variation and selection principles as opposed to, for example, rules of valid or heuristic derivation.

The second sense in which encodingism makes genuine learning (learning with emergent representation) impossible turns on the fact that learning requires error, and genuine error cannot be defined in a strict input-processing encoding system (p.49). Error *for the system*, by definition, requires some standard, from the perspective of the system itself, that can be successfully satisfied or fail to be satisfied. Learning, in turn, is a matter of how to avoid such error. Learning thus requires some sort of constructive variation of the system organization so long as errors are encountered (p.49).

Bickhard and Terveen point out the following important shortcoming of trying to account for error by a user or designer of the system: "User or designer provided error criteria simply import from outside the system the necessary supplements for learning to occur. These are no more a general solution to the problem of learning than a user or designer semantics is a solution to the problem of representation. Learning with designer provided error criteria is also fixed, unless further user or designer interventions occur: such a system cannot learn new kinds of error" (p.49). Thus, this cannot be a general approach or solution to learning because it requires all of the prior knowledge of what counts as error and success, and what the anticipatory combinatoric space is that supposedly contains the solution.

Bickhard and Terveen then argue that the requirement for error criteria and error signals in order for learning to occur yields two further, related problems for encoding approaches, thus filling out the second and third impacts of encodingism on learning. The first is that a *strict* encoding system will simply encode in some way or another (or fail to encode) *all* inputs, so the very distinction between

[11] There are a number of interesting attempts to account for this capacity via evolutionary processes on epistemic systems, including attempts to find simple primitives (e.g., "edge detectors" in the visual system) out of which these higher-level components are built.

feedback input and "just another input to be encoded and processed," must itself be pre-built into the system. Second, learning in any general sense is in response to error; but if a system is a strict, passive, encoding system, then it has no way to check if its encodings are in error — the system itself has no way of distinguishing error. On the other hand, if a system is to be able to get feedback, then it must be able to produce some output and track the results of such output — that is, it requires interaction with some environment in order to derive feedback so that it can search in its combinatoric space. An additional step is that, if the above is required for learning to occur, then *what* will be learned is, fundamentally, how to anticipate and control such interactions (i.e., interactive representation, which is introduced in the next section). This then entails that even with such pre-definitions, the system cannot be just an information processor.

Again, this overview is highly adumbrated. The above argument is one of the most complex in the book — my intent here is to just catch the high-points. This overview of the affects on learning in turn sets the stage for presenting the interactive framework, which is proposed to avoid the problems of encodingism and also offer a viable alternative on which to base future cognitive science and AI research. I turn now to the overview of interactivism.

8 - The proposed solution: interactivism

Interactivism is proposed to capture the core notions of representation while avoiding the incoherences of encodingism. Interactivism, however, is not motivated solely by the desire to avoid encodingism.[12] It is developed as a combination of several intellectual traditions whose roots have interesting similarities, but have not, until now, been integrated into a complete framework. The traditions include Piaget's constructivism, Gibson's ecological theory of perception and affordances, and the pragmatist approach to knowledge.[13] There is also convergence with the later Wittgenstein's (1953) "meaning as use" theory of meaning. Bickhard and Terveen, however, also take note of important divergences from aspects of these traditions. In all, the intellectual roots of interactivism are not explored fully in their book because the primary focus is on identifying encodingism in the field, its effects, and an outline of the alternative; however, extensive development can be found elsewhere.[14]

[12] Historically, in fact, Bickhard developed the basic interactive model *before* he discovered the encodingism critique.

[13] It should be noted that interactivism is convergent with the pragmatist focus on meaning and action, but that pragmatism's general model of representation is still encodingist.

[14] There are numerous pointers in Bickhard and Terveen's book for where in other sources specific issues are dealt with in more detail. For example: Bickhard 1980, 1987, 1995, Bickhard & Campbell 1992, and Campbell & Bickhard 1992 particularly develop the interactive framework with respect to language; Bickhard & Richie 1983 investigates Gibson's theory of perception and it's

It should again be noted here that Bickhard and Terveen first and foremost espouse a program of naturalism — they are not advocating abandoning attempts to capture intentionality, representationality and other mental phenomena within a naturalistic framework (p.9). (This point is important because in any foundational critique it is important to be clear about what it is you are rejecting categorically, and that which you are not.) By naturalism they mean, understanding phenomena (in this case, representational phenomena) in natural terms as science has come to understand other phenomena in the world: "just as life, fire, magnetism, heat and other once strange phenomena are now understood at least in principle as parts of the overall natural order, so also will mind be understood as emergent in the overall natural world" (p.150). In fact, it is in their attempts to naturalize representation that they have come to reject encodingism and adopt and develop interactivism.

In this section, I will focus on presenting an outline of interactivism from the perspective of avoiding encodingism. I will pose many of the following aspects of the model without justification; again, I leave such clarification up to Bickhard and Terveen. The main point of this section is to simply present what the proposed alternative view of representation is, and show enough of it to make it worth further investigation.

As was concluded in Section 4, encodingism is fundamentally incoherent. In particular, the result of such incoherence is that the premise of encodingism — that representational content is ultimately *constituted in* or *provided by* (or even, *the same as*) epistemic contact — is false. It is here that interactivism makes its break with encodingism: interactivism avoids the incoherences of encodingism by making a fundamental split between what constitutes representational content and epistemic contact.

Note, however, that leaving representational content (what it is in the representation that makes it *about* what it represents) *unrelated* to epistemic contact (whatever makes the functioning of a representation appropriate to the current situation in the environment) would be a mistake. This would leave us with a new dilemma: we wouldn't have any epistemic access to the world via our representations and our content would constitute all there is of our world; representation would loose its connection with the world (this would place us in a strange place, somewhere between solipsism and idealism). We still want our representations to be about the world while also not having the representations constitute all that there is of the world. Interactivism thus explains their relation so that the notion of representing is kept intact.

---

relations to interactivism; Campbell & Bickhard 1986, and Bickhard 1988 and 1992 examine Piaget's constructivism and develop the interactive model in development; and another important paper is Bickhard 1993, which focusses on the move from rejection of encodingism to adoption of interactivism.

The next move that interactivism makes, related to the fundamental rejection of encodings, is to choose a particular kind of ontology for representations — one that allows for emergence: a process ontology. This is in contrast to encodingism's basic substance ontology. For an understanding of emergence in most cases of natural science, a shift is required from a substance ontology to a process ontology of the phenomena involved. For example, historically there has been a shift between the conception of life as a vital fluid (a substance) to life as a form of open system (a process model). Within a substance ontology, the fundamental substances themselves cannot emerge. For example, the Greeks' four fundamental substances of earth, air, fire, and water could not themselves emerge, but had to be in existence from the beginning. In substance ontologies, the basic substances serve as primitives which cannot themselves be derived from within the ontology itself (p.22).

Process models, on the other hand, allow for such emergence. How such emergence occurs, however, requires an explanation — such an account is not given for free (the emergence of life, for example, *can* be explained in a process model such as considering an open system dynamics, but *how* this happens is still an open question). Some examples of systems which are best described in terms of ontologies of process include: closed system stable processes, such as atoms and molecules, and open system stable processes, such as flames (and life). Bickhard and Terveen point out that these ontologies *do* involve instantiation in material terms, but not in terms of forms of material types (p.167). For example, a flame is not just the molecules and atoms constituting it — the same material engaged in different interactions would not be a flame, and differing material substrates are in fact involved at each moment in the flame. The fact that such phenomena and their explanations in terms of process ontologies exist is unproblematic; thus, moving representation to a process ontology is not *a priori* unrealistic.

The next consideration to make is determining the nature of interactive representation in a process ontology. While encodingism focusses on the *elements* of representation (as per a substance ontology), interactivism requires a shift to a view of representation as being a *functional aspect of certain sorts of system processing*. By a *functional* model of representation it is meant that interactivism presents an explanation of representat*ion* (*how* something represents or is representing), rather than representat*ions* as things (although from the functional perspective we can make reference to the functional organization of a system which would participate in representing, as I will do below).

"In its broadest sense, the only function that a representation could serve internal to a system is to select, differentiate, the system's further internal activities. This is the basic locus of representational function, but two additional logical necessities are required" (p.58). The first is the possibility of error. This arises because the functional differentiation of the system activities must be in some sense epistemically related to some environment being represented — that is, there has to be some sense in which the system's functional representing of the environment could possibly be wrong. If a representation was never wrong, that would mean that the system directly knows the environment. This claim does seem to serve as its own reductio, as any epistemic system clearly does not have such omniscient access.[15] Representations constitute the system's model of the world, and for the system to be able to tell the difference between its model of the world and the actual world, it has to have some way of detecting that its model (its representation) could be wrong.

The possibility of error, in turn, requires that the system be able to act and interact with it's environment. This is required because without such action (without an output), the system has no capacity to determine the entailments of what it thinks the world is like based on its representation of the world. A strictly passive information processor has no way of checking such representations against the world.

The second logical necessity is that this error must be for the system itself. This arises because those differentiations must in some sense constitute at least implicit predications that could be wrong *from the perspective of the system itself.*[16] It should be clear from these necessities that interactivism is a proponent of the strong version of the "representation for the system" feature described back in Section 3. This second necessity requires some notion of *goal-directedness* — some reason for why a failure to be correct about the world has entailments for the system itself. This gives the system its own internal criteria for setting correctness — error for the system itself. A goal in this sense is some internal system state which is to be satisfied; much more goes into this argument and how to qualify what a goal is, but I will not present this here.

With the stipulation of action and interaction, in combination with the requirement that the error be from the perspective of the system, an action can then be intuitively thought of as being like an empirical test for the system itself: "I think the world is such that I can do ***X***. If I am correct (i.e., if my representation which posits that the world allows for ***X*** is correct), then I will in fact be able to do ***X***. But until I actual try to do it, I will not know if the world does allow me to do ***X***. If I can't do , then I have to revise my representation of the world." This sketch sets the foundation for how the system can shape and modify its representation through interaction with the environment. This, in turn, shows how the function of representation closes the circle from internal system function of what an action should produce, to action, and then to the system's perspective of the outcome of the action. And this is where interactivism gets it's name — *inter*activism. Figure 1 is an incomplete sketch of this cycle (the particulars of Figure 1 will be filled in what follows):

[15] Note that this is not to be confused with the surface similarity of Gibson's "direct perception" — for one, perception is not the same as knowledge.

[16] As Bickhard and Terveen note, the criteria of just simply being wrong is too weak as it allows any observer semantics to determine such "wrongness" and thus yields a semantics for that observer, but not for the system itself.

[FIGURE 1]

Figure 1: The Cycle of *Inter*action

The outline thus far presents the basic intuition of the interactive view of representation as "anticipation of interactive potentialities" (that is, anticipation of potential actions and interactions). Accounting for how these "anticipations" are implemented in the functional organization of a system, however, requires more discussion, and is the task which I turn to now. In doing so, I will fill in some more of the details concerning the particulars of interactive representation and how it can be considered representation: i.e., what constitutes epistemic contact, what constitutes representational content, their relation, and how representational content emerges.

First I will present how Bickhard and Terveen explain epistemic contact. Consider a system or subsystem in interaction with an environment. The course of the interaction will depend in part upon the organization of the system itself, but in part it will also depend upon the environment being interacted with. Thus, differing environments may yield differing flows of interaction, and differing environments will likewise leave that (sub)system in differing final internal states or conditions when the interaction is "completed." These possible internal final states will serve to *differentiate* possible environments. For example, environments that yield internal final state **S13** will be differentiated from those environments which yield internal final state **S120**. These possible final states, in turn, will *implicitly define* the class of environments that would yield that state if in fact encountered in an interaction. This differentiation and implicit definition is what constitutes *epistemic contact*.

While this differentiation and implicit definition by a final state constitutes epistemic contact (reaching of the final state is, in part, contingent on the environment), the final state itself does not indicate anything at all about its implicitly defined environments — except that they would yield that final state. Again, this is where interactivism diverges from encodingism: "A possible final state will be in *factual* correspondence with one of its implicitly defined environments whenever that state is in fact reached as a final state, but the state per se contains no information about what that correspondence is with — the relationship to the corresponding class of environments is purely implicit. *Thus there is no semantic information, no representational content, available that could make that final state an encoding*" (p.60).[17]

These possible final states constitute a basic representational *function* without themselves bearing any representational content (nothing is represented about the implicitly defined class of environments except that it is different from

[17] Note, however, that (passive) differentiations (e.g., transductions) are precisely what are standardly called encoded representations.

the other differentiated classes). As Bickhard and Terveen point out, this seemingly small point is what makes interactivism invulnerable to the incoherence problem that encodingism lands in: "In particular, an interactive differentiating final state does not require that what is being represented be already known in order for it to be represented. It is precisely that requirement for encodings that yields the incoherence of foundational encodings" (p.61).

The next task is to specify how interactivism accounts for representational content, without providing it from the observer perspective as encodingism requires. In its most concise form, *representational content* is defined as *indications of potential further interactions*. This picture is made more clear by adding a little more detail to Figure 1, producing Figure 2. Now, consider a system which has some internal state that it is trying to reach — a goal state. This goal state serves as part of the mechanism for "choosing"[18] what action to take next.

[FIGURE 2]

Figure 2: Representational content as *indications of potential further interactions*

Next, consider that the system is currently organized such that if it is in a certain state, say $\boldsymbol{n}$, then a variety of possible future interactions are indicated — in Figure 2, these indications are labeled $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\chi}$. These indications are pointers to possible actions (to be precise, an action is a pointer to an action generating subsystem: e.g., the process for the of tensing a muscle) that could be carried out. Along with the actions are also indicated the internal system states which should result given the action. In Figure 2, the functional indication $\boldsymbol{\alpha}$ indicates that if the action is made, then the internal system state $\boldsymbol{m}$ is expected to result. This internal state $\boldsymbol{m}$, in turn, functionally indicates further possible interactions, with associated action generation and expected outcomes ($\boldsymbol{\alpha_1}$, $\boldsymbol{\alpha_2}$, $\boldsymbol{\alpha_3}$) — these are the "further interactions." In this way, the indicator $\boldsymbol{\alpha}$ indicates potentials for further interaction ($\boldsymbol{\alpha_1}$, $\boldsymbol{\alpha_2}$, $\boldsymbol{\alpha_3}$ likewise each indicate sets of new further potential interactions, and so on).

These networks of indications from given states in certain given conditions constitute potential system processing. Selection is then based on what the current goal is, and how that goal might be reached if a certain indication of future potential interaction is "chosen." In this way, the system can use the differentiations in these indicated final states to differentiate the system's own internal goal-directed processing. For example, if the system in Figure 2 is trying to

[18] The word "chooses" is quoted to highlight that it is *not* intended to be taken in the sense of *necessarily* requiring some homunculus which then "decides." Rather, a choice could instead be a matter of what path the system *will* take when certain conditions hold (perhaps set by the goal state and other possible conditions). They are choices only in the sense that when we remove considerations of what particular conditions currently hold and look to see what the different kinds of possible actions are that could be taken in all kinds of different conditions, we see that there are multiple possible paths — and as will become clear shortly, these paths are indexed according kinds of conditions which make those paths appropriate. "Choices" could also, of course, be stochastic.

reach a particular goal state, then in certain conditions, choosing the action indicated by $\boldsymbol{\alpha}$ might be the path to take, while under other conditions the action indicated by $\boldsymbol{\beta}$ might be more appropriate (the conditions warrant choice of different potential future interactions).

These processing selection dependencies constitute representational content about the differentiated environmental classes. Thus, in the example from Figure 2, if I am in internal state $\boldsymbol{n}$, and, therefore, in the environmental state implicitly defined by $\boldsymbol{n}$, then that environmental state is (implicitly) predicated (by the internal indicator — the dashed line $\boldsymbol{\alpha}$ leaving state $\boldsymbol{n}$) to have the interactive properties that would yield internal state $\boldsymbol{m}$ (and, therefore, the environmental state implicitly defined by $\boldsymbol{m}$) if I should engage in interaction $\boldsymbol{\alpha}$. (Similarly, being in internal state $\boldsymbol{n}$ is also predicated to have interactive properties that would yield other states if I engaged in interaction $\boldsymbol{\beta}$ or $\boldsymbol{\chi}$.) Here, the representational contents are the possible selection-of-further-processing uses that can be made of the differentiating final states (e.g., the expected state $\boldsymbol{m}$) of a chosen interaction. Conversely, the final states and their indicators indicate the further interactive properties appropriate to whatever selections of further interaction might be made on the basis of those final states. (pp.61-62)

Although epistemic contact plays a different representational function than representational content, the two are necessarily related, and this relation is part of what constitutes the capacity for the system to be able to modify its current functional organization: if an indication outcome is not met — if what was expected to happen doesn't — then the system has available to it the information that an error has occurred. For example, in Figure 2, if outcome state $\boldsymbol{o}$ is not the same as (or close enough to) the expected internal system state $\boldsymbol{m}$, then the system may ratify its indication links. In this way, the state $\boldsymbol{m}$ has a truth value for the system (it is true if outcome state $\boldsymbol{o}$ matches $\boldsymbol{m}$, and false if they don't match), and this truth value is contingent on the actual environment. So representational contents, through interaction, depend upon the outcomes which are a result of what the environment is in fact like — but the outcomes, for the system, are in terms relative to what was expected.

This possibility for error, in turn, can be used for functionally distinguishing when learning should take place. For example, if the expected state $\boldsymbol{m}$ does not match the outcome $\boldsymbol{o}$, then the system may change its internal organization links so that in the future when in internal state $\boldsymbol{n}$, amongst the possible future interactions indicated will be a link to state $\boldsymbol{o}$ (and whatever possible future interaction links follow from it). This would constitute new representational content. A number of possible learning mechanisms could be posited now that the necessary condition of having the possibility of error which is system detectable has been satisfied.

Now an account of the emergence of representational content is possible. First, note that a model of the emergence of a functional process (like representation) must be *independent* of issues of representation because function is logically prior to representation; the emergence of representation is then modeled *within* that framework of functional process (p.93). Interactivism thus satisfies this constraint because, as is clear from above, all discussion of the functional composition of a system has been done without assuming the existence of representational components; rather, the components are then built out of these system functions. And this is what gives interactivism the capacity to account for the emergence of representation out of what is non-representational (the components of system functional organization).

The emergence is found in the potential for construction of new possible indications linked with new implicit definitions (new environmental differentiations) — that is, a new functional link can be made between some internal system state and a part of the system for producing an action[19], along with its associated expected outcome state. This possible construction is afforded precisely because of the split between representational content (which is now defined as internal indications of future interactions) and epistemic contact (which is defined as the environmental differentiation class that the internal differentiation-state implicitly defines). Such new indications can be constructed out of initially non-representational, functional components, unlike the case with encodings. Possible combinations of what is implicitly defined and what interactions are afforded are constructed by contributions from the system *and* from the environment — in no way is there the need for already representational "primitives."

"Legitimate" encodings, from the interactive perspective, are now seen as the following: what assigns an encoding's representational content is a property of the functional usage of the encoding by the system — it is a property of the system knowing what the encoding is supposed to represent — *and not a property of the encoding element itself*. In this sense, interactivism is a more fundamental form of representation than encodings in that it is possible to construct *derivative* encodings on an interactive, functional representation base, but not the other way around: "...[interactivism] provides an account of the 'ground' or 'foundation' for representational content that encodingism cannot" (pp.56-57). Thus, interactivism has the potential to account for all of the phenomena which involve derivative encoding representations.

19 Note that Bickhard and Terveen do not require that the internal subsystems for the generation of "action" be literal bodily actions which affect the environment — they may be movement of perceptual apparatus to orient the system to particular sensory stimuli (e.g., saccading the eye in a direction to see a visual object more clearly). This kind of action doesn't *affect* the environment itself (at least not much). The actions may even be more "internal" to the system; for example, waiting for something to happen, or doing a mental activity like mentally imaging an event, etc. These are not discussed in detail in the book (but are elsewhere — e.g., Bickhard & Richie 1983), but it is made clear by Bickhard and Terveen that all of these cases constitute "actions" and could be part of interactions. (Internal or "mental" "interactions" will clearly be different in status and components from external interactions.)

Bickhard and Terveen present a nice summary of the three core aspects of interactive representations (p.92):

1) **Epistemic Contact** - Interactions with an environment terminate in one of two or more possible internal final states, thus implicitly differentiating the environment with respect to those possible final states. This is the *epistemic contact* aspect of representation — the manner in which interactive representations make contact with particular environments.

2) **Functional Aspect** - Internal states or indicators, generally constructed with respect to dependencies on such final states, influence further system processing — this is the *functional* aspect of representation and is the only role representations can play within a system.

3) **Representational Content** - Through influencing goal-directed interaction, which either succeeds or fails in achieving its goals, *representational content* emerges in the organization and functioning of a system as falsifiable implicit interactive predications about the environment — representational content has truth value that is fallibly determinable by the system itself, not just by an observer.

This outline which I have presented is only the account of how interactive representations can capture the minimal requirements for being representations and how they avoid the problems of encodingism (the deepest being an account of representational content emergence). A number of issues are certainly left open, however. How can these representations scale to account of more abstract concepts, like object permanence and numbers? What are the processes that maintain the large webs of indications of potential action? What constitutes reasoning based on existing interactive representations as functional indications? Bickhard and Terveen touch on some of these issues in there book, but it is clear that much more work needs to be done to fill these gaps in. Nonetheless, interactive representations certainly have been established as a viable alternative kind of representation and Bickhard and Terveen make a good case that the interactive framework will be able to account for the cognitive phenomena which involve representation.

Throughout, but particularly near the end of the book, Bickhard and Terveen explore the entailments of interactivism for system architectures. An issue which I have not mentioned but which they devote some time to is the issue of timing. Bickhard and Terveen argue for the necessity of an account of timing in representation, and this has entailments for the kinds of architectures that can instantiate genuine interactive representations. Namely, traditional modelling approaches within the computational paradigm may be insufficient (Turing computation, strictly speaking, makes no account of timing, only sequence). In general, Bickhard and Terveen look towards dynamic systems approaches in autonomous agents as promising avenues for accounts of how interactive representation is manifested in a system's functional makeup. This is particularly interesting to note as dynamic systems research (which includes the newer situated cognition movement in cognitive science) by in large eschews trying to make accounts of representation — rather than eradicating representation, dynamic systems may be our only hope of saving representation. And interactivism may provide the crucial synthesis.

The following is a summary of strengths of interactivism compared to encodingism, as well as some other general aspects of Bickhard and Terveen's discussion which make interactivism attractive:

- Avoids the encodingism incoherence problems
- Satisfying account of representational emergence
- Naturalized account of representational content *for the system* (further system activity), which is not dependent upon an external observer
- Able to naturally capture learning - and a more powerful learning than encodingism's combinatorial spaces.
- Convergence with much of the recent developments in the study of the fundamental nature of language
- Wide convergence with many other programmatic developments:
  - (i) Gibson (similar to Gibson's critique of encodingist perception theories; adoption of affordances),
  - (ii) Piaget (similar to Piaget's critique of "copy" theories of representation; adoption of constructivism),
  - (iii) Peirce (pragmatic view of meaning and knowledge),
  - (iv) Later Wittgenstein (rejection of early Wittgenstein "picture" model; adoption of meaning as use),
  - (v) Agrees with most of situated cognition and dynamic systems research — these two approaches should be particularly interested in interactivism because here is a critique which is greatly convergent with theirs, *and* it offers a new paradigm for handling representation that is amenable to the situated and embodied notions).
  - (vi) There is also some interesting convergence with hermeneutics and Heidegger, but in a way that might be more palatable to analytically minded scientists: namely, interactivism emphatically denies taking theoretical critiques of hermeneutics to the extreme of linguistic idealism.

To sum up, it seems the deep theoretical result of the identification of encodingism and the move to interactivism is the following. Going back to my original two criteria, representational *aboutness* and representation *for the system,* it might be said that what Bickhard and Terveen see as being encodingist is that these two criteria are somehow satisfied by one-in-the-same feature of a mechanism: thus what satisfies the account of epistemic contact is also the account of representational content. Interactivism, on the other hand, makes a fundamental distinction between these two: there's the function of epistemic contact, and there's the function of representational content, and these have an important relationship, but they are crucially *not the same thing* nor satisfied by the same mechanism.

9 - A note on connectionism and PDP

In contemporary cognitive science and AI, no foundational-level discussion of representation would be complete without addressing the connectionist and PDP paradigms. Bickhard and Terveen, in fact, devote an entire section later in their book (Chapter 13) to discussing the strengths and weaknesses of the these approaches and how they are affected by encodingism. I have waited until now to discuss them because Bickhard and Terveen's main criticisms of these approaches are most clearly seen against the backdrop of the reaction of interactivism to encodingism as I've outlined above.

As Bickhard and Terveen point out, connectionism is exciting because, unlike other approaches within encodingism, it allows for the emergence of environmental differentiators: "PDP systems are, in effect, models of the emergence of logical transducers: transducers of input categories into activation patterns." (p.294). The main problem, however, is that connectionism in general, like other approaches within encodingism, treats such differentiations as representations of the environment. In general, the focus in connectionism is on the dynamics of the connectionist system which allow it to "settle" on differentiations (categorizations), but not of how representation for the system itself is to be accounted for.

Among a number of other issues which Bickhard and Terveen raise about connectionism is the argument that there is still no account of error *for the system,* or the role of content for future system function. Also, most "learning" in the connectionist paradigm is guided by a "teacher" — even most nets that "settle on their own" do so solely based on criteria set by the researcher. All of these are passive learners. And connectionism offers no model of how the representational content could be available to the system to guide its own learning. Bickhard and Terveen suggest how connectionist types of architectures might be modified to include interactive properties (e.g., how the connectionist system's output may be taken as indications of what the system would expect to result if certain actions were taken — thus making the connectionist system part of an interactive system).

10 - Take-home messages from the book

The book's three main themes, the critique of encodingism, the arguments that encodingism is pervasive in cognitive science and AI, and the proposal of interactivism to solve the problems of encodingism, all raise many questions. The book certainly uncovers new fertile ground to be tilled in philosophy and cognitive science. In this section I consider four questions whose answers I believe highlight some further important take-home messages of the book which I have not yet addressed.

(1) *If encodingism has so many problems and has been dominant for such a long time, then what is the appeal that has kept it alive?*

Bickhard and Terveen offer three main reasons (pp.51-52). First, all external representations are, in general, encodings: paintings, maps, statues, blueprints, computer codes, etc. It is apparent that mental representation cannot be *identical* to these, but it is not so apparent how fundamentally different it is. Not recognizing the fundamental difference leads to a search for the right variation or form of these kinds of representations that is the same as mental representation.

Second, all of these common external representation examples are *structural* representations; that is, they are structures of objects, properties or events. These are most easily captured in a substance ontology. Bickhard and Terveen argue that substance ontologies typically develop first, not only in the developmental history of the individual child, but also in the development of theories of phenomena (e.g., in theories of the nature of living systems, it was originally proposed that life was a product of some vital fluid; now living systems are thought of in terms of open system processes). Eventually these substance models come to be replaced by process models. Investigations of representation have "simply" not yet made that shift.

Third, encodingism itself is quite suggestive of many different permutations and approaches, and its fundamental problems seem to suggest possible answers within encodingism: a vast and intricate maze of red herrings. Thus, there are many versions of encodingism to explore, many problems to be solved, and many potential fixes for each problem, which in turn produce new problems with new potential fixes.

I would also add a fourth reason for why encodingism is appealing: encodingism allows us to make an "immediate jump" to high-level representations involving complex concepts and linguistic entities, like cars, people, political situations, etc. People don't want to hear about encodingism being false, or interactivism being true because it means we have an even longer way to go before we can capture these complexities. What has to now be asked is: Does this all therefore mean that we *have* to work from ground zero and up, or can we still do good science at higher-levels, at least approximating the phenomena we wish to explain? I believe this is an open question who's answer will require a lot of investigation.

(2) *Who is potentially affected by the encodingism critique?*

Anyone dealing with explanations of representational phenomena (which, in terms of cognitive systems, includes pretty much everything)

(3) *Who is not affected?*

For some of the practical tasks of AI, solving theoretical problems of representation is not as important. These practical tasks tend to be for cases where the systems are best construed as tools which serve as enhancements for ultimately human-supervised work. For example, expert systems are only affected by representational issues to the extent that their problem task requires solving representational problems. Usually expert systems do *not* have robust learning, thus in these cases the issues of representational content emergence are not crucial. Also, the conditions in which an expert system is to appear and for which it is designed are often situations in which most of the representational problems are handled by the creator of the system — so there is not as much concern for the epistemology of the system itself; the human handles this.

(4) *Give up computational modelling altogether?*

One of the results of Bickhard and Terveen's critique is that core notions of computational modelling are found to be problematic. Namely, Turing computation does not capture properties of timing and Tarskian semantics cannot serve as a foundational theory of representation. Does this then mean that all computational cognitive modelling is therefore not useful? I believe that answer is, and should be *no* — and Bickhard also agrees. Although computational models may have some shortcomings in capturing all of the properties of interactive representations, there are still many aspects of interactivism that can be fruitfully explored within computational models. In fact, there are several current approaches within computational modelling which could serve as a foundation for studies of interactivism: for example, artificial life models, in which systems are modeled as agents which interact with a simulated environment. Also, computational models are still the most flexible and readily available kinds of modeling mediums available. And while interactivism in general proposes a move to robotic architectures, many of the control mechanisms of such architectures are at least preliminarily designed within computational models. Granting that issues such as timing are important, it seems that computational modelling still has a lot to offer and should not be thrown out entirely.

11 - Conclusion

Bickhard and Terveen offer a radical new view of representation and picture of the state of cognitive science and AI. It turns out that some of our basic beliefs about how to model intelligence, how a system could represent its world, may be quite different from what we had thought. For a system to represent — *really* represent *on its own* — it has to be in the world and interact with it, and build its own model of the world based on those interactions. The danger of not taking Bickhard and Terveen seriously is to remain at an impasse: to not transcend the need of *already* intelligent, representing beings to act as representational surrogates for every model or attempt at explanation — meanwhile, the representational power of these already representing beings is still left unexplained. Such non-interactive models, by themselves, are hopelessly incomplete and handicapped because they lack the basic capacity to know their world. What they lack must be filled by a real interactive representors. An account of such an interactive representors constitutes a new generation of cognitive science and AI models.

Even if the arguments characterizing encodingism, the impasse that is alleged to follow from it, the claims of what positions do or do not assume encodingism, and arguments for how to avoid encodingism, do not persuade you, this book is still important to read. It offers a new model of representation that combines and augments several important intellectual traditions and is also certain to be accepted by several of the new approaches growing in cognitive science (e.g., interactivism finds a natural home in situated cognition, autonomous agents research, and dynamic systems approaches). I believe it will be important to understand the arguments in this book and the interactive framework it recommends; and it is a must read for any contemporary cognitive science and AI researcher who wishes to stay abreast of cutting edge research in semantics and representation.

References

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

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* (Orlando: Academic Press), pp.39-65.

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

Bickhard, M. H. (1992). How Does the Environment Affect the Person? In L. T. Winegar & J. Valsiner (Eds.) *Children's Development within Social Contexts: Metatheory and Theory* (Hillsdale, NJ: Lawrence Earlbaum Associates), pp.63-92.

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

Bickhard, M. H. (1995). Intrinsic Constraints on Language: Grammar and Hermeneutics. *Journal of Pragmatics*, **23**: 541-554.

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 Gibson's Theory of Perception*. New York: Praeger Publishers.

Block, N. (1986). Advertisement for a Semantics for Psychology. In P. French, T. Uehling, Jr. & H. Wettstein (eds.) *Studies in the Philosophy of Mind*, Vol. 10, *Midwest Studies in Philosophy* (Minneapolis: University of Minnesota Press)

Brooks, R. A. (1991). Intelligence Without Representation. *Artificial Intelligence*, **47**(1-3): 139-159.

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

Campbell, R. L., Bickhard, M. H. (1992). Clearing the Ground: Foundational Questions Once Again. *Journal of Pragmatics*, **17**(5/6): 557-602.

Chalmers, D. J., French, R. M. & Hofstadter, D. R. (1992). High-level perception, representation, and analogy. *Journal of Experimental and Theoretical Artificial Intelligence*, **4**: 185-211.

Clancey, W. J. (1993). Situated Action: A Neuropsychological Interpretation Response to Vera and Simon. *Cognitive Science*, **17**(1): 87-116.

Clark, A. (1993). *Associative Engines.* Cambridge: The MIT Press.

Cummins, R. (1996). *Representations, Targets, and Attitudes.* Cambridge: The MIT Press.

Dennett, D. C. (1987). *The Intentional Stance*. Cambridge: The MIT Press.

Dretske, F. I. (1981). *Knowledge and the Flow of Information*. Cambridge: The MIT Press.

Dretske, F. I. (1986). Misrepresentation. In R. Bogdan (Ed.) *Belief.* (Oxford: Oxford University Press), pp.17-36.

Fodor, J. A. & Pylyshyn, Z. (1988). Connectionism and Cognitive Architecture: A Critical Analysis. In S. Pinker & J Mehler (Eds.) *Connections and Symbols* (Cambridge: The MIT Press), pp.3-71.

Fodor, J. A. (1994). *The Elm and the Expert*. Cambridge: The MIT Press.

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* (Hilsdale, NJ: Lawrence Earlbaum Associates), pp.67-82.

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

Harnad, S. (1990). The Symbol Grounding Problem. *Physica D*, **42**: 335-346.

Haugeland, J. (1985). *Artificial Intelligence.* Cambridge: The MIT Press.

Lloyd, D. (1989). *Simple Minds.* Cambridge: The MIT Press.

Millikan, R. (1984). *Language, Thought, and Other Biological Categories.* Cambridge: The MIT Press.

Millikan, R. (1993). *White Queen Psychology and Other Essays for Alice*. Cambridge: The MIT Press.

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

Searle, J. (1980). Minds, Brains, and Programs. *Behavioral and Brain Sciences*, **3**: 417-424.

Smolensky, P. (1988). On The Proper Treatment of Connectionism. *Behavioral and Brain Sciences*, **11**: 1-74.

Thelen, E. & Smith, L. B. (1994). *A Dynamic Systems Approach to the Development of Cognition*. Cambridge: The MIT Press.

van Gelder, T. J. (1991). What is the "D" in "PDP"? A Survey of the Concept of Distribution. In W. Ramsey, S. P. Stich & D. E. Rumelhart (Eds.), *Philosophy and Connectionist Theory* (Hillsdale, NJ: Lawrence Earlbaum Associates), pp.33-59.

van Gelder, T. J. & Port, R. (Eds.) (1996). *Mind as Motion: Dynamics, Behavior, and Cognition*. Cambridge: The MIT Press.

Wittgenstein, L. (1953). *Philosophical Investigations*. Trans. G. E. M. Anscombe. Oxford: Basil Blackwell.