ELSEVIER

Journal of Cognitive Systems Research 1 (2000) 65–75

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

Information and representation in autonomous agents

Action editor: Ron Sun

Mark H. Bickhard*

Cognitive Science, 17 Memorial Drive East, Lehigh University, Bethlehem, PA 18015, USA

Accepted 13 August 1999

Abstract

Information and representation are thought to be intimately related. Representation, in fact, is commonly considered to be a special kind of information. It must be a *special* kind, because otherwise all of the myriad instances of informational relationships in the universe would be representational — some restrictions must be placed on informational relationships in order to refine the vast set into those that are truly representational. I will argue that information in this general sense is important to genuine agents, but that it is a blind alley with regard to the attempt to understand representation. On the other hand, I will also argue that a different, quite non-standard, form of information is central to genuine representation. First, I turn to some of the reasons why information as usually considered is the wrong category for understanding representation; second, to an alternative model of representation — one that is naturally emergent in autonomous agents, and that does involve information, but not in standard form; and third, I return to standard notions of informational relationships and show what they are in fact useful for.

Information and representation are thought to be intimately related. Representation, in fact, is commonly considered to be a special kind of information. It must be a *special* kind, because otherwise all of the myriad instances of informational relationships in the universe would be representational — some restrictions must be placed on informational relationships in order to refine the vast set into those that are truly representational (Smith, 1987, 1995). Perhaps the informational relationship must be causal in origin, or perhaps it must be an instance of a nomological relationship (Fodor, 1987, 1990a,b, 1998; Levine & Bickhard, 1999). Perhaps even broader informational relationships will do if some special history obtains (Dretske, 1981, 1988). Perhaps some sort of structural relationship must hold as an aspect of the informational relationship, such as the isomorphism requirement of the Physical Symbol System Hypothesis (Newell, 1980; Vera & Simon, 1993). Perhaps some special training is required to have established the informational relationship, such as for connectionist nets (McClelland & Rumelhart, 1986; Rumelhart, 1989; Rumelhart & McClelland, 1986).

I will argue that information in this general sense is important to genuine agents, but that it is a blind alley with regard to the attempt to understand representation. On the other hand, I will also argue that a different, quite non-standard, form of information is central to genuine representation. First, I turn to some of the reasons why information as usually considered is the wrong category for understanding representation; second, to an alternative

*URL: http://www.lehigh.edu/~mhb0/mhb0.html.
E-mail address: mhb0@lehigh.edu (M.H. Bickhard)

1389-0417/00/$ – see front matter
PII: S1389-0417(99)00007-8

model of representation — one that is naturally emergent in autonomous agents, and that does involve information, but not in any standard form; and third, I return to standard notions of informational relationships and show what they are in fact useful for.

1. Representation as information

The central mystery about representation focuses on representational content. Content is that which specifies for the system that 'has' the representation what it is supposed to be a representation of. Content yields 'aboutness': X represents Y involves X being about Y. If the content specification is correct, the representation is true, while if the content is not correct, the representation is false. Thus, content also yields representational truth value. Accounting for content has proven to be the downfall of virtually all proffered models of representation.

The content problem is not simple. Many purported solutions have been developed over millennia of attempts, and a vast array of subsidiary problems has emerged in the course of such investigations. Furthermore, new problems are still being discovered. Nevertheless, the assumption is that the basic approach — that representation must be some sort of informational or correspondence relationship — *has* to be correct. After all, what is the alternative? Accordingly, efforts are focused on trying to find the specific details that will avoid the multifarious problems and refine informational correspondence into genuine representation.

The claim that I will be making is that there is in fact an alternative; the myriad problems that afflict standard approaches are symptomatic of a fundamental incoherence in those approaches, not simply of difficulties that are yet to be overcome. Informational correspondence is *not* the correct framework within which to understand or model representation, or to build machines that have genuine representations. To assume that it is the correct approach is to encounter, either explicitly or implicitly, some of the vast number of impossibilities that are thereby created. It is to founder on a foundational impasse, often enough an impasse that is not even recognized, or is recognized only dimly and with mistaken diagnosis of its nature and origin (Bickhard, 1996b; Bickhard & Terveen, 1995).

A full survey of such problems is a book-length undertaking, so I will here only provide some illustrative examples. The first is a direct attack on the ability of informational correspondence models to account for the possibility of error — for the possibility that the content of the representation is incorrect. Consider: if the crucial informational relationship exists, then whatever that relationship is a relationship with also exists, so the representation is true; on the other hand, if that crucial informational relationship does not exist, then the representation doesn't exist either, and so it cannot be false. There are three distinct possibilities that must be accounted for in any model of representation: (1) the representation exists and is correct, (2) the representation exists and is incorrect, and (3) no representation exists — but informational models (and all other kinds of correspondence models) can distinguish only two possibilities: (1) the crucial relationship exists, or (2) it does not exist. Accounting for the very possibility of the second case, for the possibility of representational error, has proven to be one (of many) of the central ongoing failures. Attempts have been made, and are still being made, with rejoinders and counter-arguments in abundance, but there is no acceptable proposal on the table.

One such attempt is that of Fodor's notion of asymmetric dependency (Fodor, 1990a; Loewer & Rey, 1991). Asymmetric dependency attempts to capture the intuition that representational error is parasitic on representational success. The basic idea is that the possibility of *correct* evocations of a representation — say, an evocation of a representation of a cow, COW, by an actual cow — will be *independent* of the possibility of *incorrect* evocations of that representation — say an evocation of COW by a horse on a dark night — but that the error possibility, in contrast, *is* dependent on the possibility of the no-error possibility. That is, evoking the representation COW by a horse on a dark night is dependent on the possibility of COW being evoked by cows, but the possibility of COW being evoked by cows is not dependent on the possibility of its being evoked by horses on dark nights. So, the possibility of error is dependent on the possibility of

success, but asymmetrically so: the dependency is not reciprocated.

The asymmetric dependency that Fodor makes constitutive of representation seems roughly correct for error in general, but it is not specific to representation at all. Consider a neurotransmitter docking on a receptor of a neuron. It triggers various activities in the receiving cell that carry information about the transmitter, and about whatever causally preceded the release of the transmitter. But there is at best a functional story to be told here, not a representational story.

Furthermore, now consider a poison molecule which mimics the transmitter molecule as it docks on the same receptors. Similar processes are activated inside the receiving cell, again yielding informational relationships, though now with the poison and its history. But the possibility of these poisonous, parasitic relationships is dependent upon those created by the neurotransmitter, and asymmetrically so. The dependence is not reciprocated. The poison could not work unless the transmitter did work, but the transmitter could work just fine even if the poison did not. So we have parasitic informational relationships that are asymmetrically dependent on normal informational relationships, but there is still no representation in this story. Fodor's model does not distinguish representational error from functional error (Bickhard, 1993; Levine & Bickhard, 1999).

A strengthened focus on the issue of representational error asks how it is that any such representational error, assuming it were to exist, could be detected by the system. After all, to check if my representation of this desk (presumably) in front of me is correct I need to compare the desk representation with the world, with the desk if there is one, to confirm or deny the specifications of the DESK representational content. But the only way to do that is to re-invoke my desk representation: my only epistemic access to the world is via such representations. This is circular, and provides no genuine check at all. Variants of this argument have been at the core of skepticism for millennia, and have never been defeated. At best, various counter-arguments have purported to show that the conclusion of radical skepticism — that we really know nothing (Sanches, 1988) — must be false, though even in this limited sense there are no consensually accepted such arguments. It might be argued, for example, that the radical skeptical conclusion cannot be correct because it presupposes the very world, or language, that it questions (Rescher, 1980). But none of these arguments have diagnosed what is wrong with the radical skeptical argument.

Yet it must be not only the *conclusion* of radical skepticism that is false. The argument per se would, if correct, destroy all of artificial intelligence and cognitive science — the possibility of error is of fundamental scientific importance as well as philosophical importance, not to mention being essential to autonomous agents. In particular, without the possibility of system-detectable error, there could be no error-guided action and no error-guided learning. It would not be possible to account for representational origins if there were no ability to correct for error: all representation would have to be present and correct from the beginning. Clearly this is not so and cannot be so. Therefore there *must* be something wrong with the radical skeptical argument per se, not only with its conclusions.

This point about correcting for error makes contact with another fundamental problem. Standard approaches have no way of accounting for the *emergence* of representational content (Bickhard, 1993, 1997c, 1998a; Bickhard & Campbell, in press). At best they account for determining which representations are true and to be believed and which are false and not to be believed (side-stepping the impossibility of detecting any such errors in the first place). The representations that are thus confirmed or falsified must already be present. Contemporary models of learning at best account for such issues of confirmation, not for the origin of the representations at issue.

One conclusion from such considerations is that representational content must be innate. If all of the basic representational contents necessary for understanding our worlds are present genetically, then learning and development can be 'simply' a matter of determining which combinations of these innate atoms are correct and which are not. This is the basic outline of Fodor's innatism argument (Bickhard, 1991a; Fodor, 1981). Unfortunately (or fortunately), the problem concerning the emergent origins of content is logical. It does not depend on any particular models of learning or development. If content cannot emerge, then it cannot emerge in evolution

any more than it can emerge in learning or development. Conversely, if there is some way in which content *can* emerge in evolution, then there is an absent argument about why that sort of emergence is not possible in learning and development. The specifics of Fodor's argument, therefore, are inconsistent, even though the basic premise that we have no model of representational emergence is correct.

But if representation cannot emerge at all, then it cannot exist. There were no representations at the moment of the Big Bang. Therefore, if representational emergence is impossible, then representation is impossible. Conversely, there *are* representations now; therefore representation *has* to have emerged. Therefore, any model of representation that makes such emergence impossible is refuted. In particular, informational correspondence models are refuted.

The basic point here is simply that informational correspondence, even should it exist, does not announce on its sleeve *that* it exists, or what it is in correspondence with. Some state or event in a brain or machine that *is* in informational correspondence with something in the world must in addition have content about what that correspondence is with in order to function as a representation for that system — in order to be a representation for that system. Any such correspondence, for example, with this desk, will also be in correspondence (informational, and causal) with the activities the retina, with the light processes, with the quantum processes in the surface of the desk, with the desk last week, with the manufacture of the desk, with the pumping of the oil out of which the desk was manufactured, with the growth and decay of the plants that yielded the oil, with the fusion processes in the sun that stimulated that growth, and so on all the way to the beginning of time, not to mention all the unbounded branches of such informational correspondences. Which one of these relationships is supposed to be the representational one? There are attempts to answer this question, too (e.g., Smith, 1987), but, again, none that work (Bickhard & Terveen, 1995).

I will end the sampling of such problems with informational correspondence approaches for modeling representation. They are myriad and multifarious; they are of millennial age without refutation. The only reason that they have not been taken as conclusive is that they seem to yield such obviously false and unacceptable conclusions about nothing being known. But this worry presupposes that there is no alternative way to understand and model representation, one that might not be subject to these problems. To the contrary, however, there is such an alternative.

2. Interactive representation

Correspondence and informational approaches to representation have been dominant in Western history since the ancient Greeks. The general alternative framework within which I will outline a model has been available only for about a century. This alternative is pragmatism (Joas, 1993; Rosenthal, 1983, 1986). Correspondence approaches stem from taking consciousness as the locus for understanding mind, and a passive input processing receptive conception of vision as the model or metaphor for consciousness. Pragmatism suggests that action and interaction are the best framework for understanding mind, including representation.

Ultimately, of course, we want to understand both action and consciousness. The issue at hand, however, is which is the better overall framework within which to begin. There are a number of general considerations to take into account here, but I will not focus on these more general issues now. Just to illustrate, note that the classical approaches tended (and still tend) to assume a fundamental breach between humans and other animals. That assumption is not viable since Darwin, and pragmatism, in fact, makes for a much stronger connection between human and animal mental processes.

Pragmatism has offered and stimulated more than one model of representation itself. Peirce had such a model, as did Piaget, and others. I do not think that any of these models got the details right, though I am in full agreement with their choice of framework. There is no time or space here to analyze these alternative pragmatist — action-based — models of representation (Bickhard, 1988; Bickhard & Campbell, 1989; Bickhard & Terveen, 1995). Instead, I will turn to an outline of the model that I offer. It is called the interactive model of representation.

Any system, natural or artificial, that interacts with

the world faces a problem of action selection — what to do next. In simple cases, this can be solved with some sort of simple triggering relationship: if this input is received and the system is in that state, then do action X. Such triggering suffices only if there is sufficient reliability in the world that the action triggered is always the correct one, or if there is so little cost if it is not the right one or is unsuccessful that nothing is lost. These conditions can hold in sufficiently simple cases, but for more complex organisms and agents, there can easily be novel combinations of inputs and internal conditions that require a more sophisticated process of selection.

Here is a general solution to this problem. If the system can internally indicate, in appropriate conditions, that some interaction, say, X, is possible, and that it can be anticipated to yield internal outcome Q if engaged in, then we have the ground for action selection. In particular, if such an indication is available and Q is in some sense desirable, then X might be a good selection to make.

There are some delicate issues here that need to be unpacked. I will only adumbrate here how to handle them. First, if the indicated outcome is external to the system or organism, then we have the problem of detecting and representing that outcome. This is a source of circularity if our task is to model representation. That is why the critical indications must be of *internal* outcomes — they can be functionally detected, and do not circularly invoke representation.

Second, if the criteria of 'desirability' are given by system goals, and if goals are themselves representational, then we have another source of circularity. The processes of selection of action, however, do not have to involve explicit goals (Bickhard, 1993; Bickhard & Terveen, 1995), and, even if explicit goals are involved, they do not have to involve representations of the goal states. They can, for example, be 'simple' functional tests on internal conditions, such as blood sugar threshold, that switch one way if the criterion is in fact met and a different way if the criterion is in fact not met. That criterion does not have to be represented at all in order for such switching to work.

Indications of potential interactions and their internal outcomes, then, do not necessarily invoke representation, and therefore do not create a circularity as a possible framework for understanding representation. They do provide a solution to the action selection problem. For current purposes, their most important characteristic is that they provide a solution to the representation emergence problem. I turn now to explicating some of the crucial senses of that emergence.

First, indicating the potentiality of an interaction and outcome in this situation is predicating something of the current environment. It is predicating that this interaction outcome is in fact possible in this environment. It is predicating that this environment has properties that are sufficient for that interaction to yield the indicated outcomes if engaged in. And it is a predication that might be false — the indicated outcomes might well not occur if the interaction is engaged in. If engaged in, the predication can be checked by the system itself. If the indicated internal outcomes are not reached, that can be functionally detected, and such detection can influence further activity in the system.

That is, such indications constitute predications that have truth value, and truth value which is in principle detectable by the system itself. This is the fundamental form of the emergence of representation.

The content here is the set or organization of properties that would make the predication true if those properties were present in the environment. Note that this foundational form of content is implicit, not explicit. The indication per se does not specify what those properties are. It only provides a way to determine if they hold or not. This inherent implicitness is quite unlike standard models, and it is a source of great power in this model. (Such implicitness, for example, dissolves the frame problems — see Bickhard & Terveen, 1995.)

Interactive indications provide a way to select actions; they provide error feedback if the actions do not yield anticipated outcomes; they provide emergent truth value; and they provide emergent content. They account for the emergence of representation. But they do not look much like standard representations, such as of objects and numbers. There are many challenges that can be brought to this general interactive model concerning its adequacy to all forms and properties of representation. I will briefly address three of them: how to

account for explicit content; object representations; and representations of abstractions such as numbers.

An indication of the potentiality of an interaction, say X, is a predication that *this* is an X-type environment. Suppose the interaction is engaged in and the anticipated outcome is obtained. At this point, some further indications may hold, perhaps of Y. It may be that the indication system is organized such that *all* X-type environments are indicated to be (also) Y-type environments. The contents that make something an X-type or Y-type environment remain implicit, but the indicative relationship that all X environments are Y environments is explicit in the relationship between the potential indications. Here is a primitive version of explicit content. And again it is one that is in principle false and falsifiable by the system itself.

This point also provides the clue to the representation of objects. Indicative relationships can iterate: encountering X environments can indicate Y environments, which might in turn indicate Z environments. And they can branch: encountering an X environment might indicate Q, R, and S possibilities. With such iterating and branching as resources for constructing more complex representations, vast and complex webs of indications become possible.

Some sub-webs of representational indications can have additional special properties. In particular, they can be relatively closed, reachable, and invariant. Consider a toy block. It offers many possible interactions, ranging from visual scans to manipulations to dropping and throwing and chewing, and so on. Many of these possibilities require intermediate interactions in order to bring them into immediate accessibility — the block may have to be turned, for example, before a certain visual scan of the pattern on that side of the block is immediately possible. But any of these interactions will indicate the possibility of all of them, and engaging in any of them neither creates nor destroys any of the others as possibilities. That is, the web of such interactions is closed — none of the interactions takes the system out of the sub-web — and it is internally reachable — any point in the web can be reached via the appropriate intermediate interactions from any other point in the web.

Furthermore, that sub-web is invariant with respect to an important class of other interactions, such as putting in the toy box, leaving in another room, hiding behind or inside something, and so on. It is not invariant, however, with respect to all possible interactions, such as burning or crushing. The representation of a toy block, then, can be accomplished by such a web of interactive indications with the appropriate properties of closure, reachability, and invariance. This is an essentially Piagetian model of object representation (Piaget, 1954).

I suggest a similarly Piaget-inspired model of the representation of numbers. A property that an interactive sub-system might manifest is that of ordinality — do this once, or twice, or three times. Such simple counts can be important in control systems. If a second level of interactive system could interact with and represent properties of a first level that interacted with and represented an external environment, then that first level would be in effect the environment for the second level. Such a second level could represent properties instantiated in the first level, such as those of ordinality, and many others. A hierarchy of such levels of representationality provides a rich resource for the representation of abstractions, and yields many interesting predictions about how systems — children, for example — could access such higher levels and what new possibilities would thereby be opened up (Campbell & Bickhard, 1986).

The interactive model, then, does offer the possibility of modeling explicit and complex and abstract representations. It also has implications for many other cognitive phenomena that I will not address here, such as perception (Bickhard & Richie, 1983; see Gibson, 1966, 1977, 1979), language (Bickhard, 1980, 1987, 1992a, 1995; Bickhard & Campbell, 1992; Campbell & Bickhard, 1992a), rationality (Bickhard, 1991b; Bickhard, in preparation), and others (Bickhard, 1992b, 1997a,b, 1998b; Bickhard & Campbell, 1996a,b; Campbell & Bickhard, 1992b).

And it does not fall to the problems of standard informational correspondence approaches. It arises naturally in any complex interactive system, natural or artificial. It easily models the emergence of representation and representational content. Error, and even system-detectable error, is a natural phenomenon. There is no regress of correspondences with a consequent mystery concerning which is the representational one and how that manages to be so. And so on.

Interactivism is at least a candidate for modeling

the basic nature of representation. There is an alternative to the multiple aporia of informational correspondence approaches.

2.1. Information

How is information involved in this approach? In standard models, the informational relationship is backward-oriented in time and external to the system. It is with some external locus of causal connection, for example, in the past. It might be with the reflection of light off a surface that then entered the eyes. The representational problem in this scenario is to represent what produced, or at least preceded or informationally accompanied, the production of the ostensible representation in the animal or machine. We have seen that there are very good reasons to conclude that that is not a plausible model for representation.

The indicative relationships in interactive representations *are* informational relationships — they provide information to the system concerning what processes are accessible. But there are several differences from standard approaches. Interactive representation is future-looking. It is concerned with future potentialities of interaction and internal outcome. So the informational relationships are future-oriented, not past-oriented. And they are oriented internally to the system, not to the environment.

Content, in this view, emerges in the dynamic presuppositions of those informational indications, in the properties that would support those indications of future potentialities. Content is not constituted in the informational relationships per se.

A third difference is subtle, but of critical importance. Interactive representation is constituted in certain kinds of internal information about future potentialities, but the truth value of such a representation is not given in such informational relationships. The truth value is constituted in whether or not the environment supports those informational relationships, not in the functional informational relationships per se. Therefore, the interactive model is not subject to the basic aporia that if the informational relationship exists, then the representation is correct, while if the informational relationship does not exist, then the representation does not exist and therefore cannot be incorrect. A future-oriented functional indicative informational relationship can well exist and yet its relationship to the environment can nevertheless be false.

3. Information and representation

So information is crucially involved in interactive representation, but not in standard ways. Is there any role for information ‘about’ the environment in the standard sense, and, if so, what is it? There would certainly seem to be strong support for the notion that such informational relationships exist — sensory ‘encoding,’ for example, is ubiquitous. What are these doing if they do not constitute representation?

Consider an interaction X that might or might not yield internal outcome Q. If it is engaged in and does yield Q, then the system is in an X(−Q) environment, with whatever properties that involves. Strictly as a matter of *fact*, arriving at such an outcome creates an informational relationship between the outcome internal to the system and whatever the properties are that supported that outcome. This is information in the strict sense of covariation, not semantics. For, regardless of that information, there is no representation posited in this model about those properties. They remain implicit. Nevertheless, this is precisely the point at which standard approaches want to conclude that there is somehow representation, not merely information, about those properties or states of affairs in the environment that the informational relationship is with.

To see this, consider a simple version of such an interactive system — one with no outputs. Such a (sub-)system will simply process inputs, perhaps in complex ways, but will not interact. Such a passive interactive system, say X, can still have indicated internal outcomes and will still differentiate X environments from those that are not X environments, just not with as much potential full power as a full interactive system. As before, the interactive model does not attribute any representational content to such a differentiation. But, in the version of sensory ‘encoding’ or transduction, such passively generated, input processing generated, informational relationships are the paradigm of standard models of representation (Bickhard, 1993; Bickhard & Terveen, 1995; Fodor, 1990a). They are precisely what the eyes or ears are supposed to generate (Carlson, 1986).

Input processing, with its attendant factual informational relationships, then, are present in both standard approaches and in the interactive approach. The fundamental difference between the two is that standard approaches want those differentiations and attendant informational relationships to somehow constitute representations with full representational content, and so on. That seems to be impossible, and the interactive model does not depend on any such interpretation. There are no 'sense data,' or any cousin thereof, in the interactive model (Bickhard & Richie, 1983).

What, then, are those sensory differentiations and informational relationships useful for if they do not constitute representations? More broadly, what is any environmental differentiation useful for? The answer is already implicit in discussions above. Interactive representation is constituted in indications of further interactive potentialities, but under what conditions should such indications be set up? They should be set up if some prior interaction has yielded an outcome that the system takes to indicate the future potentiality.

So, encountering an X environment can indicate a Y environment, and that indication is itself representational. But what about the original encounter with the X environment? That encounter differentiates X environments from others, and thereby creates a factual informational relationship with whatever constituted this environment as an X environment, but it does not *represent* what that might be. Representation occurs in the use of such differentiations for creating future indications.

Informational relationships with the world, then, are of critical importance in this model. They are necessary for the appropriate invocation of the representational indications. They are required for the activities and representations of the system to be appropriately modulated by the environment. A random or otherwise disconnected setting up of action anticipations would definitely not do. But those differentiations are not representations. Whatever constitutes this environment as an X environment is not represented, and error is not even definable at this level of analysis: the interaction differentiates whatever it is that it differentiates. The truth value issue arises in the question of what further indications can be set up on the basis of such an X environment differentiation.

More broadly, the interactive model requires precisely the kinds of processing that are commonly dubbed representational in standard approaches, thus yielding the standard maze of problems, the standard fundamental impasse. But the interactive model does not attribute representationality to that processing or to its results, thus avoiding those problems and that impasse. It claims a crucial but different function for such differentiating processes, one that does not generate innumerable aporia.

4. Autonomous agents

The critiques offered are quite general. They apply to any version of informational correspondence model of representation. It does not matter for these purposes if the model of one of isomorphic correspondence relationships, as in the Physical Symbol System Hypothesis (Newell, 1980; Vera & Simon, 1993), or trained correspondences with activation vectors, as in connectionist models (McClelland & Rumelhart, 1986; Rumelhart, 1989; Rumelhart & McClelland, 1986), or transduced or causal or nomological relationships, as in many philosophical models (Fodor, 1987, 1990a,b, 1998), or if they are the products of various genealogical histories, as in many contemporary models of function and of representation as function (Godfrey-Smith, 1994; Millikan, 1984, 1993; see Bickhard, 1998a). Such relationships are crucial to the functioning of interactive systems, including the representational functioning, but they are not *constitutive* of representation. One aspect of this point is that transductions and connectionist nets can be important parts of interactive systems, but they do not in themselves constitute representations for those systems.

For all such correspondence models, the representational nature of purported representations does not depend on actions or interactions. Actions may follow on, and may make use of, representations in these models, but representation per se does not require that. In these models, then, representation can occur in perfectly passive systems. And standardly, action is either absent or secondary in the development and logic of such models.

Not so in the interactive model. If representation is emergent in interactive systems, as is the case for

any pragmatist model of representation, then representation is metaphysically impossible in passive systems. Conversely, any system that is interactive must solve the action selection problem, and, in all but the simplest cases, the natural solution to that problem is also the point of emergence of primitive representation. The initial emergence of primitive representation, therefore, as well as the evolution of more complex representation, is naturally accounted for in this model.

This encounter with the necessary emergence of interactive representation by virtue of encountering the action selection problem holds as strongly for artificial agents as it does for natural agents. We should find research in artificial agents, autonomous robots, encountering this problem and solving with emergent interactive representation, *whether or not* the researchers are themselves aware of it or even of this set of issues. The action selection problem and its obvious solution is simply not avoidable.

This is precisely what we find. Dissatisfaction with standard conceptions of representation is widespread, especially in autonomous agent research and dynamic system approaches (Maes, 1990), but this commonly produces a claimed rejection of *all* representation (Beer, 1990, 1995a,b; Brooks, 1991a,b; Port & Van Gelder, 1995). Nevertheless, these very research programs also produce robots that involve indications of, anticipations of, interactive potentialities (Nehmzow & Smithers, 1991, 1992; Stein, 1994; see Bickhard, 1978). There are also recognitions of the necessity of a non-standard account of representation, though most of these advocate approaches that ultimately do not escape the basic problems (Clark, 1997; Pfeifer & Verschure, 1992a,b; Prem, 1995; for discussions of, for example, CYC, Harnad, Searle, SOAR, Varela, and many others, see Bickhard & Terveen, 1995). There are also, however, recognitions of the involvement of genuine interactive representation, and of the power for design and understanding of that recognition (Bickhard, 1996a, 1997c, 1998a,b,c,d; Cherian & Troxell, 1995a,b; Christensen et al., in preparation; Hooker, 1995, 1996).

The basic point, however, is that interactive representation is being investigated of necessity in research on autonomous agents, sometimes knowingly and often unknowingly. Research is much improved if this encounter is engaged in knowingly, not to mention if the failure of standard informational approaches is recognized.

5. Conclusions

Standard informational correspondence approaches to representation have failed for millennia, and they continue to fail. They are fundamentally incoherent — they *presume* representational content, but claim to *account* for it, and *cannot* account for it.

An alternative interactive approach to representation, part of the general pragmatist approach, has been available only for about a century, and is therefore much less explored. Nevertheless, it promises to avoid the foundational impasses generated by standard models of representation. Interactive representation emerges naturally in any complex interactive system, natural or artificial. Information is crucially involved, but not in standard ways. Primitive versions initiate a natural evolutionary or constructive trajectory that can yield complex representations, such as of objects and abstractions. The interactive model of representation is a strong candidate for capturing the basic nature of representation and, therefore, for guiding research involving representation in all areas of cognitive studies.

Acknowledgements

Thanks are due to the Henry R. Luce Foundation for support during the preparation of this paper, and to Cliff Hooker, Norm Melchert, and Wayne Christensen for very useful discussions of these issues.

References

Beer, R. D. (1990). Intelligence as adaptive behavior, Academic Press, New York.

Beer, R. D. (1995a). Computational and dynamical languages for autonomous agents. In: Port, R., & van Gelder, T. J. (Eds.), Mind as motion: Dynamics, behavior, and cognition, MIT Press, Cambridge, MA.

Beer, R. D. (1995b). A dynamical systems perspective on agent–environment interaction. *Artificial Intelligence 73*(1/2), 173.

Bickhard, M. H. (1978). The nature of developmental stages. *Human Development 21*, 217–233.

Bickhard, M. H. (1980). Cognition, convention, and communication, Praeger, New York.

Bickhard, M. H. (1987). The social nature of the functional nature of language. In: Hickmann, M. (Ed.), Social and functional approaches to language and thought, Academic Press, New York.

Bickhard, M. H. (1988). Piaget on variation and selection models: Structuralism, logical necessity and interactivism. *Human Development 31*, 274–312.

Bickhard, M. H. (1991a). The import of Fodor's anticonstructivist arguments. In: Steffe, L. (Ed.), Epistemological foundations of mathematical experience, Springer, New York.

Bickhard, M. H. (1991b). A pre-logical model of rationality. In: Steffe, L. (Ed.), Epistemological foundations of mathematical experience, Springer, New York.

Bickhard, M. H. (1992a). How does the environment affect the person? In: Winegar, L. T., & Valsiner, J. (Eds.), Children's development within social context: Metatheory and theory, Erlbaum, Hillsdale, NJ.

Bickhard, M. H. (1992b). Commentary on the age 4 transition. *Human Development 35*(3), 182–192.

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. (1996a). The emergence of representation in autonomous embodied agents. In: *Papers from the 1996 AAAI Fall symposium on embodied cognition and action*, Chair: Maja Mataric. 9–11 November 1996, Cambridge, MA, MIT. Technical Report FS-96-02. Menlo Park, CA.: AAAI Press.

Bickhard, M. H. (1996b). Troubles with computationalism. In: O'Donohue, W., & Kitchener, R. F. (Eds.), The philosophy of psychology, Sage, London.

Bickhard, M. H. (1997a). Cognitive representation in the brain. In: Dulbecco, R. (Ed.), Encyclopedia of human biology, Academic Press, New York.

Bickhard, M. H. (1997b). Is cognition an autonomous subsystem? In: Nualláin, S. Ó., McKevitt, P., & MacAogáin, E. (Eds.), Two sciences of mind: Readings in cognitive science and consciousness, John Benjamins, Amsterdam.

Bickhard, M. H. (1997c). Emergence of representation in autonomous agents. *Cybernetics and systems: Special issue on epistemological aspects of embodied artificial intelligence 28*(6), 489–498.

Bickhard, M.H. (1998a). A process model of the emergence of representation. In: G.L. Farre, T. Oksala (Eds.), *Emergence, complexity, hierarchy, organization: Selected and edited papers from the ECHO III Conference*. Acta Polytechnica Scandinavica, Mathematics, Computing and Management in Engineering Series No. 91, 3–7 August 1998, Espoo, Finland, 263–270.

Bickhard, M. H. (1998b). Levels of representationality. *Journal of experimental and theoretical artificial intelligence 10*(2), 179–215.

Bickhard, M.H. (1998c). Robots and representations. In: R. Pfeifer, B. Blumberg, J.-A. Meyer, S.W. Wilson (Eds.), *From animals to animats 5: Proceedings of the Fifth International Conference on Simulation of Adaptive Behavior*. Zurich, Switzerland, 17–21 August 1998. Cambridge, MA: MIT.

Bickhard, M.H. (1998d). Whither representation? In: M.A. Gernsbacher, S.J. Derry (Eds.), *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*, University of Wisconsin-Madison, 1–4 August 1998, Hillsdale, NJ: Erlbaum, 150–155.

Bickhard, M.H. (in preparation). Critical principles: On the negative side of rationality. See http://www.lehigh.edu/~mhb0/critical.html.

Bickhard, M.H.,with Campbell, Donald T. (in press). Emergence. In: P.B. Andersen, N.O. Finnemann, C. Emmeche, & P.V. Christiansen (Eds.), *Emergence and downward causation*. Aarhus, DK: University of Aarhus Press. See http://www.lehigh.edu/~mhb0/emergence.html.

Bickhard, M. H., & Campbell, R. L. (1989). Interactivism and genetic epistemology. *Archives de Psychologie 57*(221), 99–121.

Bickhard, M. H., & Campbell, R. L. (1992). Some foundational questions concerning language studies: With a focus on categorial grammars and model theoretic possible worlds semantics. *Journal of Pragmatics 17*(5/6), 401–433.

Bickhard, M. H., & Campbell, R. L. (1996a). Developmental aspects of expertise: Rationality and generalization. *Journal of Experimental and Theoretical Artificial Intelligence 8*(3/4), 399–417.

Bickhard, M. H., & Campbell, R. L. (1996b). Topologies of learning and development. *New Ideas in Psychology 14*(2), 111–156.

Bickhard, M. H., & Richie, D. M. (1983). On the nature of representation: A case study of James J. Gibson's theory of perception, Praeger, New York.

Bickhard, M. H., & Terveen, L. (1995). Foundational issues in artificial intelligence and cognitive science: Impasse and solution, Elsevier, Amsterdam.

Brooks, R. A. (1991a). Intelligence without representation. *Artificial Intelligence 47*(1–3), 139–159.

Brooks, R. A. (1991b). New approaches to robotics. *Science 253*(5025), 1227–1232.

Campbell, R. L., & Bickhard, M. H. (1986). Knowing levels and developmental stages, Karger, Basel.

Campbell, R. L., & Bickhard, M. H. (1992a). Clearing the ground: Foundational questions once again. *Journal of Pragmatics 17*(5/6), 557–602.

Campbell, R. L., & Bickhard, M. H. (1992b). Types of constraints on development: An interactivist approach. *Developmental Review 12*(3), 311–338.

Carlson, N. R. (1986). Physiology of behavior, Allyn and Bacon, Boston.

Cherian, S., & Troxell, W. O. (1995a). Intelligent behavior in machines emerging from a collection of interactive control structures. *Computational Intelligence 11*(4), 565–592.

Cherian, S., Troxell, W.O. (1995b). Interactivism: A functional model of representation for behavior-based systems. In: Morán, F., Moreno, A., Merelo, J.J., Chacón, P., *Advances in artificial life: Proceedings of the Third European Conference on Artificial Life*, Granada, Spain, 691–703. Berlin: Springer.

Christensen, W.D., Collier, J.D., Hooker, C.A. (in preparation). Autonomy, adaptiveness, anticipation: Towards autonomy-theoretic foundations for life and intelligence in complex adaptive self-organising systems.

Clark, A. (1997). Being there, MIT Press, Cambridge, MA.

Dretske, F. I. (1981). Knowledge and the flow of information, MIT Press, Cambridge, MA.

Dretske, F. I. (1988). Explaining behavior, MIT Press, Cambridge, MA.

Fodor, J. A. (1981). The present status of the innateness controversy. RePresentations, MIT Press, Cambridge, MA.

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

Fodor, J. A. (1990a). A theory of content, MIT Press, Cambridge, MA.

Fodor, J. A. (1990b). Information and representation. In: Hanson, P. P. (Ed.), Information, language, and cognition, University of British Columbia Press, Vancouver, BC.

Fodor, J. A. (1998). Concepts: Where cognitive science went wrong, Oxford University Press, Oxford.

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

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

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

Godfrey-Smith, P. (1994). A modern history theory of functions. *Nous 28*(3), 344–362.

Hooker, C. A. (1995). Reason, regulation and realism: Toward a naturalistic, regulatory systems theory of reason, State University of New York Press, Albany, NY.

Hooker, C. A. (1996). Toward a naturalised cognitive science. In: Kitchener, R., & O'Donohue, W (Eds.), Psychology and philosophy, Sage, London.

Joas, H. (1993). American pragmatism and German thought: A history of misunderstandings. In: Joas, H. (Ed.), Pragmatism and social theory, University of Chicago Press, Chicago.

Levine, A., & Bickhard, M. H. (1999). Concepts: Where Fodor went wrong. *Philosophical Psychology 12*(1), 5–23.

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

Maes, P. (1990). Designing autonomous agents, MIT Press, Cambridge, MA.

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

Millikan, R. G. (1984). Language, thought, and other biological categories, MIT Press, Cambridge, MA.

Millikan, R. G. (1993). White queen psychology and other essays for Alice, MIT Press, Cambridge, MA.

Nehmzow, U., & Smithers, T. (1991). Mapbuilding using self-organizing networks in 'really useful robots'. In: Meyer, J. -A., & Wilson, S. W. (Eds.), From animals to animats, MIT Press, Cambridge, MA.

Nehmzow, U., & Smithers, T. (1992). Using motor actions for location recognition. In: Varela, F. J., & Bourgine, P. (Eds.), Toward a practice of autonomous systems, MIT Press, Cambridge, MA, pp. 96–104.

Newell, A. (1980). Physical symbol systems. *Cognitive Science 4*, 135–183.

Pfeifer, R., & Verschure, P. (1992a). Beyond rationalism: Symbols, patterns and behavior. *Connection Science 4*(3/4), 313–325.

Pfeifer, R., & Verschure, P. (1992b). Distributed adaptive control: A paradigm for designing autonomous agents. In: Varela, F. J., & Bourgine, P. (Eds.), Toward a practice of autonomous systems, MIT Press, Cambridge, MA, pp. 21–30.

Piaget, J. (1954). The construction of reality in the child, Basic, New York.

Port, R., & Van Gelder, T. J. (1995). Mind as motion: Dynamics, behavior, and cognition, MIT Press, Cambridge, MA.

Prem, E. (1995). Grounding and the entailment structure in robots and artificial life. In: Morán, F., Moreno, A., Merelo, J.J., Chacón, P. *Advances in artificial life: Proceedings of the Third European Conference on Artificial Life*, Granada, Spain, 39–51. Berlin: Springer.

Rescher, N. (1980). Scepticism, Rowman and Littlefield, Totowa, NJ.

Rosenthal, S. B. (1983). Meaning as habit: Some systematic implications of Peirce's pragmatism. In: Freeman, E. (Ed.), The relevance of Charles Peirce, Monist, La Salle, IL, pp. 312–327.

Rosenthal, S. B. (1986). Speculative pragmatism, Open Court, La Salle, IL.

Rumelhart, D. E. (1989). The architecture of mind: A connectionist approach. In: Posner, M. I. (Ed.), Foundations of cognitive science, MIT Press, Cambridge, MA, pp. 133–160.

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

Sanches, F. (1988). That nothing is known, Cambridge University Press, Cambridge.

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

Smith, B. C. (1995). On the origin of objects, MIT Press, Cambridge, MA.

Stein, L. A. (1994). Imagination and situated cognition. *Journal of Experimental and Theoretical Artificial Intelligence 6*, 393–407.

Vera, A. H., & Simon, H. A. (1993). Situated action: A symbolic interpretation. *Cognitive Science 17*(1), 7–48.