How to Fill Empty Symbols

Mark H. Bickhard

The empty symbol problem arises from a misconception of the nature of representation, and is solved within an alternative framework for understanding representation. In particular, genuine representation — for the system itself — does not emerge from a passive spectatoring of the environment, but rather from a future-oriented anticipatory aspect of the selection of actions and interactions by an active agent in that environment.

Symbolic Encodings and Their Problems

Symbolic Encoding. I begin with a diagnosis of why the empty symbol problem arises. Symbols are manipulable representational elements. Any manipulable unit *plus representational content* is a symbolic representation. The deep difficulty is in providing, or accounting for, representational content (the problem of content is not unique to symbolic approaches).

A symbol can be defined in terms of other symbols — so long as those other symbols already have content. For example, "A = X ^ Y" is a perfectly acceptable definition so long as "X" and "^" and "Y" are already representations. These symbols, in turn, might themselves have definitions: definitions can iterate, but such iteration must halt at some level of basic representations. The crucial question is how these basic symbols have content.

If the basic level, atomic, representations out of which all others are constructed or defined have content, then so will higher level representations. But nothing has yet proven satisfactory as a way of providing atomic level representations with content. They are empty of content, therefore all higher level symbols are also — the symbols are empty, and we have the empty symbol problem.

In practice, users and designers of computational systems provide content by interpreting the symbols as having content — *for those users and designers*. The provides enormous practical usefulness, but does nothing to solve the problem of modeling genuine cognitive representation, such as, presumably, the representations in those users and designers.

Correspondence. Almost universally, the attempt is made to provide content via some sort of correspondence with what is to be represented. "**X** corresponds to **Y**, and thereby represents **Y**" is the rule that is assumed, but there are many varieties of correspondence that are proposed as the crucial kind: The notion of transduction assumes that causal correspondence somehow creates content. Fodor proposes that nomological relationships are the critical form of informational correspondence to create or constitute content. The general information processing approach similarly assumes that informational correspondence is key, though is usually silent on just what kind of informational correspondence. Connectionism is a type of informational correspondence approach in which *trained* informational correspondence is assumed to provide or constitute content.

Fatal Problems with Correspondence. Unfortunately, there are multitudinous, multifarious problems with correspondence approaches to representation. Here is one: if the correspondence exists, the "representation" exists, *and* it is correct, but if the correspondence does not exist, then the "representation" does not exist, and, therefore, cannot be in error. The problem of accounting for the possibility of representational error is among the many serious problems facing correspondence notions, one that has received a great deal of attention in the last fifteen years, but is in fact of ancient provenance.

Here are some other problems. The problem of too many correspondences: any element in (causal, informational, nomological) correspondence with something will also be in correspondence with many other things — which is the representational correspondence? Activity in the brain that is in causal correspondence with a chair via light transmission, for example, will also be in causal correspondence with activities in the retina, light patterns and processes, quantum processes in the surface of the chair, the past history of the chair, the origins of the materials out of which the chair is constructed, and so on back to the Big Bang. What picks out the chair (presumably) as the crucial correspondence to constitute a representational content?

Here's another. How could the system or agent ever detect that it is in error? Any check is circular — any check simply uses the same "representation" to check as is being checked because there is no other, no independent, epistemic access to the world via which the check can take place (this is just an ancient radical skeptical argument, one that still stands).

Here's another. Factual correspondences with **Y** can provide information about **Y**, and thus represent **Y**, to someone who knows about the correspondence. But knowing about the correspondence is already representational. (In fact, it is already to represent **Y**.) We have a circularity.

And so on.

Interactive Representation

Correspondence encoding models have failed for over two millennia. The radical skeptical argument against the possibility of checking our representations for error, for example, has never been avoided. Correspondence approaches look backwards in time — they look down an input stream of informational relation. They are "spectator models" in the term of the pragmatists. They are passive — technologically sophisticated waxed slates that the world impresses itself into. They have no outputs, no interaction — or none that are important to the presumed nature of representation. I argue that, in fact, a future oriented framework — one with outputs, with action and interaction — is the proper framework for modeling representation.

Interactive Differentiation. Note first that the internal course and internal outcome of a system-environment interaction will differentiate environmental categories. Those environments that yield "this" internal outcome will be differentiated from those environments that yield "that" internal outcome. There is no representational content here, but such differentiations might nevertheless be useful. Note, however, that passive versions of such differentiations are commonly construed as transduced or processed or connectionist representations.

Anticipation and Truth Value. In particular, environmental differentiations might be useful in selecting next interactions: If an **A** type of environment is encountered, then interaction type **X** with outcome **Q** is possible. If **Q** is relevant to the system's goals, then **X** might be selected (or avoided) accordingly. Such indications anticipate the outcomes of the indicated interactions *in the differentiated environments*.

Such anticipations can be in error, and such error can be detected by the system itself: If the interaction **X** is engaged in, then internal outcome **Q** will occur — to check, engage in **X** and check to find out if **Q** occurs. This is system detectable error. This is system (not designer or user) emergent truth value.

Anticipatory action and interaction selection yields the emergence of primitive representation: system activity with truth value. More complex, more familiar, representations, such as of objects, can be constructed as complex organizations of such anticipations and the dependencies among them.

Pragmatism and Piaget. The interactive model of the emergence of representation out of action connects with the pragmatist tradition and with Piaget. Generally Piagetian accounts of object and abstract representations, such as numbers, are thereby available.

What about Content? An indication of the availability of an interaction with associated internal outcomes is a predication of the current environment: this environment is an **X**-interaction type of environment. The contents of that predication are the properties, *whatever they are*, that would in fact support the indicated interaction yielding the indicated outcomes.

The content, the properties, are *implicitly* represented — implicit in the general interactive predication. They are not explicitly encoded. This is a dynamic version of implicit definition in model theory. Such implicitness is a source of great power — for example, it dissolves the frame problems (just to introduce a little potential controversy).

Non-Empty, Contentful Symbols

Interactive implicit content can provide content to symbols; it can fill otherwise empty symbols. How? How can a symbol be defined in terms of an interactive anticipatory process?

Encodings are stand-ins. In the definition $A = X \wedge Y$, A serves as a stand-in for $X \wedge Y$. A carries the same content as $X \wedge Y$ by virtue of being an abbreviation or substitute for $X \wedge Y$. The critical point, then, is that encodings can also stand-in for *interactive* representations. Interactive representation is anticipation about the potentiality for particular interactions and associated outcomes. The contents are implicitly defined properties that would support those potentialities. An element can stand-in for such indications of potentiality in the sense of being *used* as indicating such potentiality in action and interaction selection.

Manipulable Non-Empty Symbols. When not participating in determining the flow of interaction, such elements can be manipulated — for example, perhaps in explicit inference processes that would yield further interactive indicating symbols. These symbols are manipulable, should that be useful, but they carry agent defined content that is inherent in the system's own action and interaction selection processes.

These interactive based symbols, in other words, are not empty. They carry emergent content with truth values of and by and for the system itself, not just for users or designers.

Suggested References

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

Bickhard, M. H. (1996). Troubles with Computationalism. In W. O'Donohue, R. F. Kitchener (Eds.) *The Philosophy of Psychology*. (173-183). London: Sage.

Bickhard, M. H., Campbell, R. L. (1996). 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*. New York: Praeger.

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

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.