First published in *Behavioral and Brain Sciences* 1982, *5*, 385–386. Commentary on Roitblat, H. L. (1982) The meaning of representation in animal memory (ibid. pp. 553–406).

The Logic of Representation

Do rats and pigeons act under the influence of representations? I don't know, and Roitblat has not convinced me that he does, either.

When organism o's encounter with stimulus S at the time t_i makes a difference for what o later does at time t_k, evidently o's total state at each intervening time t_j must include some condition R_j whose instantiation by o at t_j has been brought about by o's experiencing S at t_i and which, in turn, affects o's behavior at t_k. But that scarcely begins to imply that R_j "represents" S (or anything else) for o at t_j. Otherwise, any stage of any causal process, even one as simple as the chained falling of dominoes, would count as a representation of each stage that preceded it. To justify talk of behavioral "representations," it must be argued not only that the link between input and output is mediated by an internal condition of some complexity, but also that the mediator has the right *kind* of complexity both in function and in composition. Unfortunately, although intuitions about representation pervade our everyday thinking, such notions are still far too obscure to provide insights into behavioral mechanisms. Theories about these mechanisms are much to be desired—but with representation figuring therein as explicandum, not as explicans.

We can surely agree that entity R is not a representation, unless there is some other thing T that R represents, or would represent were T to exist, under some mapping principle ρ that picks $T = \rho R$ out from all other aspirant representees. (Whether R's p-relatum must always be strictly unique will not be discussed here.) But a representational system is not just any old triple $\langle R, \rho, T\rangle$ in which ρ is a function from domain $\mathbf{R}$=$\{R_i\}$, onto range $\mathbf{T} = \{T_j\}$, or else we should have to admit, for example, that telephone subscribers represent the numbers assigned to them by phone directories. Linguistic intuition demands more than that, something else that we intimate by saying that representations "depict," "symbolize," "signify," or "stand for" their objects (see any dictionary). For $\langle R, \rho, T\rangle$ to yield "depictions," complexes of **R**-units must correspond in some nontrivial fashion to certain complexes of **T**-things under element-to-element mapping ρ. And R_i symbolizes/signifies/stands for T_i only if it does so for some organism o at a time t by somehow standing proxy for T_i in o's occurrent or dispositional functioning at t. But complexes of what sorts? And proxy how?

Three paradigms of representation can be discerned within established linguistic usage. Most basic is the primary cognitive "aboutness" (intentionality à la Brentano) that characterizes the contents of mental acts such as believing, perceiving, fearing, desiring, remembering, trying, and the like. Here, element domain R comprises concepts that purportedly designate (refer to, denote, are of) concrete or abstract objects, and form compounds whose psychonomic nature is still unknown,[1] most notably "propositions" which, when truthful, depict facts.

Second, when some suitably structured external or internal stimulus S_i seems capable of evoking in organism o an intentional content C_i that signifies some real or possible entity T_i distinct from S_i, we often say that S_i represents T_i for o, especially when, as in perception, the primary elicitor of C_i is T_i itself. Language is the example par excellence; but pictures, schematics, and most other stimulus patterns that common sense takes to encode "information" also illustrate well this concept of secondary aboutness. When S_i so represents T_i for o, it stands proxy for T_i in being a means for o to think about T_i in the latter's absence. Indeed, o's percept of pattern S_i (or, if S_i is internal, S_i's embodiment in o) may well correspond structurally to the composition of o's concept C_i of T_i under an elicitational transduction of parts in such fashion that S_i can evoke C_i in o (i.e., depict T_i) even when o has never encountered T_i itself.[2] Such an S_i encodes information for o precisely to the extent that it makes available to o the propositional content of a belief.

Finally, modern mathematics formally views one relational system $\langle \mathbf{R}, \phi \rangle$ as a representation of another, $\langle \mathbf{T}, \psi \rangle$, under mapping ρ whenever (a) ϕ and ψ are binary (more generally n-ary) relations on respective domains $\mathbf{R}$ and $\mathbf{T}$, (b) ρ is a function from $\mathbf{R}$ onto $\mathbf{T}$, and (c) for all R_i, R_j in $\mathbf{R}$, $\phi(R_i, R_j)$ holds if and only if $\psi(\rho R_i, R_j)$ does. Despite superficial similarities, this is very different logically from cognitive representation, even though we have good reason to suspect that morphisms of ideation under causal mapping are important in the psychonomics of secondary aboutness. When $\langle \mathbf{R}, \phi \rangle$ is isomorphic or homo-morphic to $\langle \mathbf{T}, \psi \rangle$ under ρ, and $\langle T_i, T_j \rangle = \langle \rho R_i, \rho R_j \rangle$, the fact $\phi(R_i, R_j)$ formally represents the fact $\psi(T_i, T_j$ relative to mapping ρ but not relative to any particular organism o that this is a representation for; and $\phi(R_i, R_j)$ also simultaneously represents many

[1] No, Virginia, whatever propositions and other compound concepts may be, they surely do not much resemble the bundles of associations proposed by current semantic-network models. For that matter, neither do we have any satisfactory theories of the cognitive-aboutness relation (see Rozeboom, 1979)

[2] That is, given a logic in which notation '$F(a_1, \ldots, a_n)$' designates a pattern comprising elements $a_1, \ldots, a_n$ distinctively positioned by index within structural frame $F(\_, \ldots, \_)$, we envision a piecemeal establishment of elicitational connections $\{a_i \rightarrow b_i\}$ and $\{F_j \rightarrow G_j\}$ in consequence of which $F_j(a_1, \ldots, a_n)$ elicits $G_j(b_1, \ldots, b_n)$ as distinct, for example, from $G_j(b'_1, \ldots, b'_n)$ for permutation $\langle b'_1, \ldots, b'_n \rangle$ of $\langle b_1, \ldots, b_n \rangle$.

other facts $\{\psi'(T'_i, T'_j\}$ relative to other mappings $\{\rho' : \mathbf{R} \rightarrow \mathbf{T}'\}$.[3] Moreover, it is logically impossible for a formal representation $\phi(R_i, R_j)$ to be false, although degrees of approximate morphism can be built into more elaborate conceptions of formal representation. In contrast, were notation '$\phi(R_i, R_j)$' to denote a cognitive representation, $\phi(R_i, R_j)$ would be a pattern, not a fact; what it signified would be relative to *o* but not to an outside observer's arbitrary choice of a mapping relation, and its depiction could well be erroneous. Confusion between the formal and cognitive senses of representation is one reason why information-processing generalities in modern psychology are seldom more than inchoate metaphors.

How well do the "neural representations" conjectured by Roitblat fit into this tripartite scheme? Despite his opening declaration that this label means nothing more than behavioral mediation, his subsequent statements read more like representation intuited as secondary aboutness. No other reading makes sense in view of his pervasive descriptions of mediators as representations specifically of the organism's salient environment rather than of the many other entities to which they are even more nearly isomorphic, and his unhesitant assumption that these code information for *o* about the world.

But if that is so, then Roitblat's claim that the research he cites demonstrates representations in animals is quite unwarranted. We still know almost nothing about the specific psychonomic mechanisms that embody even the most exemplary instances of cognitive aboutness, but surely mediators abound, especially in lower animals, that are neither primary cognitions nor vehicles for their elicitation. In all likelihood these often have a functional/compositional character intermediate between the cognitive/noncognitive extremes; but to savor the nuances of their multifaceted similarities and seek a behavior-theoretic counterpart of chemistry's periodic table, we must first take pains to appreciate their larger differences.

When we can plausibly explain some particular behavioral phenomenon in terms of a computationally effective mediation theory that does not include aboutness as part of the mechanism (see Fodor 1980) and avoids magical words like "code" and "information" unless these are clearly defined for this application, then it is time to consider whether mediational processes like these may be the psychonomic basis of common sense cognitive representations.

References

Rozeboom, W. W. (1979). On behavioral theories of reference. *Philosophy of Science*, *46*, 175–203.

[3] For every relational system $\langle \mathbf{R}, \phi \rangle$ and one-one mapping ρ of $\mathbf{R}$ onto $\mathbf{T}'$, $\langle \mathbf{R}, \phi \rangle$ is isomorphic under ρ to the relational system $\langle \mathbf{T}', \psi' \rangle$ in which ψ' is the image of ϕ in $\mathbf{T}'$ under ρ'.