

MENTAL REPRESENTATION

Blackwell Companion to Epistemology, Blackwell, 1992

Robert Cummins

I. TERMINOLOGY

Representation. Contemporary philosophy of mind, following cognitive science, uses the term 'representation' to mean just about anything that can be semantically evaluated. Thus, representations (and, by extension, representing systems or devices) may be said to be true, to refer, to be true of something, to be about something, to be accurate, etc. Representation thus conceived comes in many varieties. The most familiar are pictures and three-dimensional models (e.g., statues), linguistic text (including mathematical formulas), and various hybrids of these such as diagrams, maps, graphs and tables. It is an open question in cognitive science whether mental representation falls within any of these familiar sorts.

It is equally an open question in the philosophy of representation just what varieties of representation are possible, and what distinguishes them. Two approaches are distinguishable: different kinds of representation might be distinguished by how they represent, or they might be distinguished by what kind of semantic content they have. An example of the first approach would be the claim that pictures differ from linguistic text in that pictures are similar to the things they represent, whereas linguistic texts are not. An example of the second approach would be that linguistic texts have propositions, or truth/reference conditions as contents, whereas pictures do not.

Content. "Content" has become a technical term in philosophy for whatever it is a representation has in virtue of which it represents,

i.e., in virtue of which it is semantically evaluable. Thus, a statement is sometimes said to have a proposition or truth condition as its content; a term is sometimes said to have a concept as its content. Much less is known about how to characterize the contents of non linguistic representations than is known about characterizing linguistic representations. Pictures and maps, for example, can be accurate or inaccurate, suggesting that they have accuracy conditions as statements have truth conditions, but little research has been done on the semantic characterization of non-linguistic representation. "Content" is a useful term precisely because it allows one to abstract away from questions about what semantic properties representations have; a representation's content is just whatever it is that underwrites its semantic evaluation.

II. REPRESENTATION AND THOUGHT

The representational theory of cognition. It is uncontroversial in contemporary cognitive science that cognitive processes are representation-manipulating processes. Just as one calculates by manipulating numerals - representations for numbers - so one thinks by manipulating mental representations. This idea seems nearly inevitable. What makes the difference between processes that are cognitive - solving a problem, say - and those that are not - a patellar reflex, for example - is just that cognitive processes are epistemically assessable. A solution procedure can be, justified and correct; a reflex cannot. But epistemic assessment only makes sense of things with content: no epistemic assessment without semantic evaluation. Thus, processes appear to count as cognitive only in so far as they implicate representations.

It tempting to think that thoughts - especially beliefs, desires and intentions - are the mind's representations: Aren't thoughts just those mental states that have (semantic) content? This is, no doubt, harmless enough provided we keep in mind that cognitive science may attribute to thoughts properties and contents that are foreign to

common sense. First, we should keep in mind the possibility that the contents common sense attributes to thoughts might be inexplicit contents, hence not the explicit contents of any representations. Thus, cognitive science might refuse to countenance thoughts sanctioned by common sense. Second, most of the representations hypothesized by cognitive science do not correspond to anything common sense would recognize as thoughts, for their contents are typically not accessible to consciousness, nor understandable to non-specialists. Standard psycholinguistic theory, for instance, hypothesizes the construction of representations of the syntactic structures of the utterances one hears and understands, yet we are not aware of the structures represented, and non-linguists understand specifications of such structures no better than non-chemists understand specifications of redox reactions. Thus, cognitive science may attribute thoughts where common sense would not. Finally, cognitive science may find it useful to individuate thoughts in ways foreign to common sense. (See the discussion of "internalistic" theories of content below.)

The representational theory of intentionality. The representational theory of cognition gives rise to a natural theory of intentional states such as believing, desiring and intending. According to this theory, intentional states factor into two aspects: a functional aspect that distinguishes believing from desiring and so on, and a content aspect that distinguishes beliefs from each other, desires from each other, and so on. A belief that p might be realized as a representation with the content that p and the function of serving as a premise in inference, while a desire that p might be realized as a representation with the content that p and the function of both initiating processing designed to bring it about that p and terminating such processing when a belief that p is encountered.

III. REPRESENTATION AND THE THEORY OF CONTENT.

Naturalizing content. A great deal of philosophical effort has been lavished on the attempt to naturalize content, i.e., to explain in non-semantic, non-intentional terms what it is for something to be a representation (have content), and what it is for something to have some particular content rather than some other. There appear to be only four types of theory that have been proposed: theories that ground representation in (i) similarity, (ii) covariance, (iii) functional role, and (iv) teleological theories.

Similarity theories hold that *r* represents *x* in virtue of being similar to *x*. This has seemed hopeless to most as a theory of mental representation because it appears to require that things in the brain must share properties with the things they represent: To represent a cat as furry appears to require something furry in the brain.

Covariance theories hold that *r*'s representing *x* is grounded in the fact that *r*'s occurrence covaries with that of *x*. This is most compelling when one thinks about detection systems: the firing a neural structure in the visual system is said to represent vertical orientations if its firing covaries with the occurrence of vertical lines in the visual field. Dretske (1981; 1986) and Fodor (1987; 1990) have, in different ways, attempted to promote this idea into a general theory of content.

Functional role theories hold that *r*'s representing *x* is grounded in the functional role *r* has in the representing system, i.e., on the relations imposed between *r* and other representations in the system's repertoire by specified cognitive processes. Functional role theories take their cue from such common sense ideas as that someone cannot believe that cats are furry who does not know that

cats are animals or fur is like hair. For a defense of the functional role approach, see Block (1986), Harmon (1982), and Loar (1982).

Teleological theories hold that r 's represents x if it is r 's function to indicate (i.e., covary with) x . Teleological theories differ depending on the theory of functions they import. Perhaps the most important distinction is that between historical theories of functions (Millikan (1984, Papineau (1987)), and a-historical theories (Fodor (1987; 1990), Stampe (1977), Cummins (1989)).

Theories of representational content may be classified according to whether they are atomistic or holistic and according to whether they are externalistic or internalistic. Atomistic theories take a representation's content to be something that can be specified independently of that representation's relations to other representations. What Fodor (1989) calls the crude causal theory, for example, takes a representation to be a $|cow|$ --a mental representation with the same content as the word 'cow'--if its tokens are caused by instantiations of the property of being-a-cow, and this is a condition that places no explicit constraints on how $|cow|$'s must or might relate to other representations. Holistic theories contrast with atomistic theories in taking the relations a representation bears to others to be essential to its content. Thus, functional role theories take a representation to be a $|cow|$ if it behaves like a $|cow|$ should behave in inference.

Internalist theories take the content of a representation to be a matter determined by factors internal to the system that uses it. Thus, what Block (1986) calls "short armed" functional role theories are internalist. Externalist theories take the content of a representation to be determined, in part at least, by factors external to the system that uses it. Covariance and teleological theories that invoke an historical theory of functions, take content to be determined by "external" factors. Crossing the atomist- holist distinction with the internalist-externalist distinction gives us the

following table:

	• Atomist	• Holist
Externalist	• Historical teleological • Covariance	• Long-armed functional role
Internalist	• A-historical teleological • Similarity	• Short-armed functional role

Externalist theories (sometimes called non-individualistic theories, following Burge (1979)) have the consequence that molecule for molecule identical cognitive systems might yet harbor representations with different contents (Putnam (1975), Burge (1979), Millikan (1984)). This has given rise to a controversy concerning "narrow" content. If we assume some form of externalist theory is correct, then content is, in the first instance "wide" content, i.e., determined in part by factors external to the representing system. On the other hand, it seems clear that, on plausible assumptions about how to individuate psychological capacities, internally equivalent systems must have the same psychological capacities. Hence, it would appear that wide content cannot be relevant to characterizing psychological equivalence. Since cognitive science generally assumes that content is relevant to characterizing psychological equivalence, philosophers attracted to externalist theories of content have sometimes attempted to introduce "narrow" content, i.e., an aspect or function of wide content that is equivalent in internally equivalent systems. (I am ignoring certain refinements of formulation here. See, e.g., Block (1986), for details.) The simplest such theory is Fodor's idea

(1987) that narrow content is a function from contexts (i.e., from whatever the external factors are) to wide contents.

IV. MISREPRESENTATION

Perhaps the most serious single problem facing the theory of representation is to give a satisfactory account of misrepresentation. Consider covariance theories. A paradigm case of misrepresentation is a case in which a $|cow|$ is tokened in response to a horse. But if r is tokened in response to a horse, then its tokenings don't covary with the occurrence of cows, hence it cannot be a $|cow|$. Perhaps it represents something common to cows and horses, e.g., the property of being a cow-or-horse. (This is why Fodor (1987) has dubbed the problem the disjunction problem.) But then it is not a $|cow|$, and not a misrepresentation. Again, misrepresentation of x degenerates into proper representation of something else.

Functional role theories also face an analogous problem. A case of error is a case in which Σ tokens a representation m that it uses as a $|mouse|$, in a situation in which it should have tokened something else, say s , something Σ uses as a $|shrew|$. So m was tokened by a process C that is designed to work only if the slot that Σ fills with m is something that represents shrews. But if C uses m , then m isn't used (exclusively) as a $|mouse|$ in Σ , contrary to hypothesis. It appears that, at a minimum, Σ uses m as a $|mouse\ or\ shrew|$. Hence, m represents being-a-mouse-or-shrew, and there was no error after all.

One way to deal with the kind of problem just rehearsed is to construe representational error as a mismatch between what r actually represents and what r is used to represent on some particular occasion. For, to use r as a $|mouse|$, say, is for some process that requires a $|mouse|$ in a certain role to token r in that role. That process will treat r as if it were a $|mouse|$. If it is not,

then, barring luck or a very insensitive task, the error in representation will issue in an error in performance (i.e., in the output of the process).

This approach requires a notion of representational content that is independent of how a representation is used. It is obvious that functional role theories cannot provide this, for they explain content in terms of use. But it is equally clear that covariance theories explain content in terms of use when we reflect that to say that r covaries with x is to specify one aspect of r 's functional role, hence one aspect of r 's use.

While a use-independent conception of representation has the advantage of allowing a clear conception of representational error, many find it implausible. What, other than the way a representation is used, could give it a semantic content? The only alternative to exploiting use to give a naturalistic account of content appears to be similarity. Yet similarity theories seem hopeless as theories of mental representation, as we've seen. Still, perhaps some very abstract mathematical notion of similarity can be harnessed for the job, or perhaps there is some other alternative as yet undiscovered. Teleological theories appear to have the right form to deal with misrepresentation, for they distinguish between what r indicates and what it is r 's function to indicate, thus making room for error as a mismatch between the two. A representation r misrepresents x , according to this approach, when it is r 's function to indicate x but r does not indicate x . Thus, teleological theories have the rather odd consequence that if r misrepresents x , r represents x : misrepresentation is a species of representation.

A more serious problem is that many representations do not have indication as their function. Teleological theories are therefore forced to hold that representations that do not have indication as their function are complex, the idea being that non-indicating representations derive their content from their structure together

with their indicating constituents. Teleological theories are therefore committed to a kind of verificationist reductionism: every representation must reduce to those whose function is to indicate.

Aside from general worries about semantic reductionism, this commits the teleological theory to the view that every serious scheme of mental representation must have a componential semantics. While this might be true (see Fodor and Pylyshyn (1988) for an argument), it appears to be an empirical hypothesis about the mind, and therefore should not be an analytic consequence of the theory of content.

BIBLIOGRAPHY

Block, N. 1986. "Advertisement for a Semantics for Psychology." *Midwest Studies in Philosophy* , 10:615-678

Burge, T. 1979. "Individualism and the Mental." *Midwest Studies in Philosophy* .4: 73-121.

Cummins, R. 1989. *Meaning and Mental Representation* . Cambridge, Mass.: MIT Press. A Bradford Book.

Dretske, F. 1981. *Knowledge and the Flow of Information* . Cambridge, Mass.: MIT Press. A Bradford Book

Dretske, F. 1986. "Misrepresentation." In *Belief* . R. Bogdan ed. Oxford: Oxford University Press. Dretske (1981

Fodor, J., and Pylyshyn, Z., 1988 "Connectionism and Cognitive Architecture." *Cognition*, 28: 3-71.

Fodor, J. 1990. *A Theory of Content and Other Essays* . Cambridge, Mass.: MIT Press. A Bradford Book.

Fodor, J. 1987. *Psychosemantics: the Problem of Meaning in the Philosophy of Mind*. Cambridge, Mass.: MIT Press, a Bradford Book.

Harmon, G. 1982. "Conceptual role semantics." *Notre Dame Journal of Formal Logic* 23:242-256

Loar, B. 1982. "Conceptual role and truth conditions." *Notre Dame Journal of Formal Logic*. 23:272-283

Millikan, R. 1984. *Language, Thought, and Other Biological Categories*. Cambridge, Mass.: MIT Press. A Bradford Book.

Papineau, D. 1987. *Reality and Representation*. Oxford: Basil Blackwell.

Putnam, H. 1975. "The Meaning of 'Meaning'." In K. Gunderson, ed. *Language, Mind and Knowledge*. Minnesota Studies in Philosophy of Science Series. Vol. 7.

Stampe, D. 1977. "Towards a Causal Theory of Linguistic Representation." In P. French, T. Uehling, and H. Wettstein eds. *Midwest Studies in Philosophy*. University of Minnesota Press: 42-63.