Content in Computation

Daniel Seymour
Department of Philosophy
City University of New York Graduate Center
33 West 42nd Street
New York, NY 10036
sey@cunyvms1.gc.cuny.edu

Abstract

Examining the philosophical foundations of theories in computational psychology, and cognitive science in general, is a methodology that is likely to yield strong results to problems in the philosophy of mind. One such problem is the problem of intentionality. An intentional property is semantic: it has parts which refer or are true. The problem is to explain why these properties are empirically, and hence causally, respectable. As in all special, "non-basic" sciences, an empirically respectable property has sufficient conditions for its instantiation. But specifying such conditions for the intentional properties used by computational psychology proves difficult, since apparently neither physical nor computational identity are enough. A solution is proposed by examining in some detail the computational theory of vision. A key element of this theory requires that the intentional properties attributed to representations are constrained by considering the later computational uses to which these representations must be put. This constraint is strong enough to yield sufficient conditions for a given representation to have a given intentional property. Since analogous constraints are likely to be found in other cognitive domains, the result argued for constitutes an important methodological and philosophical insight about cognitive science in general.

I

Computational psychology is a particular type of undertaking in cognitive science, and cognitive science is a particular type of science. It is, like astronomy, geology, and biology, a special science. Special sciences are unique in ways which, say, physics is not: their explanations and predictions can misfire for reasons independent of their proprietary domains and their generalizations typically require explanation in terms of some more "basic" science.

But computational psychology is unique in ways which other special sciences are not: some properties the theory mentions are *intentional*. An *intentional property* is one which is about something, say, because it has *content*—it has parts which refer or have extensions or are true. In this paper, I argue that this feature of computational psychology does not compromise its status as a special scientific theory and that its having that status may be precisely what is required in order to make some inroads into a solution to the philosophical problem of intentionality.

II

Before elaborating that problem, consider the structure of the special sciences. There is a need in these sciences to specify lower-level micro properties which are sufficient for the instantiation of higher-level macro properties. Were such higher-level properties not accounted for in this way, they would be of dubious empirical standing. To use Jerry Fodor's (1987, ch. 1) example, rivers have the macro property of eroding their outer banks (ceteris paribus). We can explain how rivers instantiate that property by specifying sufficient conditions for erosion involving the forces and momentums of flowing water, compositional and spatial properties of rivers and their banks, etc. Notice that the generalization "rivers cause erosion" can be converted into a generalization quantifying over mechanisms: there is a mechanism of forces, momentums, and other properties which makes it the case that if something is a river, then it is an eroder. Understanding the situation in terms of mechanisms is useful, for it allows us to appeal to the same nomic mechanism in cases where erosion occurs in the absence of rivers and their banks: e.g., when prevailing winds erode sand dunes. Here, too, there is a mechanism of forces, momentums, etc. which brings it about that if something is a prevailing wind blowing on a sand dune, then it is an eroder. In short, the relation between erosion and the underlying mechanism is sufficient both ways, even though the mechanism and hence the erosion may be realized in varying physical systems—rivers, sand dunes, oceans, mountains, glaciers, lakes, etc. In such cases, we take the mechanism to be sufficient for the disjunction of physical realizers and each of the disjuncts to be sufficient for the mechanism.

On the face of it, the picture is much the same in computational psychology. Take, for example, the generalization "beliefs that Fa cause beliefs that $(\exists x)(Fx)$". Parallel to the geological case, we can explain how beliefs that Fa can instantiate the property of causing beliefs that $(\exists x)(Fx)$ by specifying sufficient conditions involving the computational properties of the corresponding psychological states. Notice, too, that the generalization can be converted into one which quantifies over mechanisms: there is some computational mechanism which brings it about that if one believes that Fa, then one believes that $(\exists x)(Fx)$. In addition, computational mechanisms are typically multiply realizable, in that various hardware architectures can realize the same computational functions. The picture is summarized in Figure 1. So, one might conclude, we get the same sort of sufficiency relations in computational psychology as we do in geology. Hence, psychology turns out to be a genuine special science with empirically grounded, well-behaved properties.

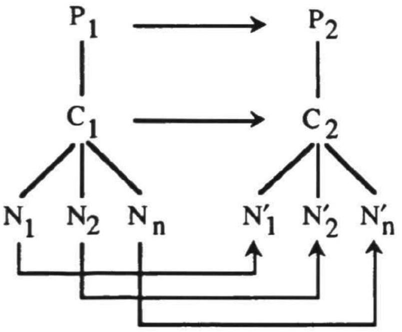

Figure 1. The structure of the special sciences. The arrows represent nomological relations between properties and the lines indicate sufficiency relations between them. In the case of psychology, the *P*'s are psychological properties, the *C*'s are computational properties, and the *N*'s are physical properties (e.g., neural or silicate). For geology, the *P*'s represent geological properties, the *C*'s represent Newtonian mechanical properties (or whatever), and the *N*'s represent physical properties (e.g., tidal or glacial).

Unfortunately, there are difficult philosophical questions arising for psychology that do not arise for geology, or for any other special science. Psychological properties are characteristically intentional—that is, they are contentful properties. If I believe that frogs rule Mars, then I have a certain property, viz., that of believing that frogs rule Mars. And that property has truth conditions: it is true if and only if frogs do indeed rule Mars. The difficulty is that various pre-theoretical intuitions about the intentional contents of psychological states apparently challenge the otherwise useful and desirable assimilation of psychology into the structure of the special sciences generally.

The problem is best illustrated by Twin Earth thought experiments proposed by Hilary Putnam (1975) and later, in a slightly different form, by Tyler Burge (1979). Putnam has us imagine a planet alike in every physical respect to Earth except that on Twin Earth there is no H_2O. On Twin Earth, the lakes, puddles, streams, and oceans are filled with a substance of a radically different chemical composition, XYZ, even though XYZ looks, tastes, and feels exactly like H_2O. Since Twin Earth is a physical duplicate of Earth in every other respect, each of us here on Earth has a physically indistinguishable Twin counterpart on Twin Earth. In such a situation, the computational structure of Twins' brains are identical, since computations are, by definition, local processes: they depend only upon solipsistic properties of the system doing the computing—the local syntactic environment, or whatever. But, the intuition goes, Twins are psychologically (i.e., intentionally) distinct in virtue of their being causally related to distinct environments. If I believe that water is wet, then my Twin believes, as we would say, that twater is wet ("twater" is an English word which refers to XYZ). Hence, computational properties, which are local, cannot be sufficient to fix the intentional properties of psychological states. To make the same point in another way, sameness of computational structure does not entail sameness of intentional structure.

In terms of the architecture mentioned above, such a result would be undesirable, for there would no longer be available similar answers to two distinct and important questions. (Q1) How do mental states acquire contentful properties? and (Q2) How do mental processes preserve content? Computational processes are defined over symbols (i.e., formally specified representations) and the rules driving the processes manipulate those symbols solely with respect to their syntactic properties. So, if the syntax is principled and rule-governed, and the interpretation of the symbolic representations is systematic and coherent, then the semantics will take care of itself. That is, the rules are specified relative to the syntactic properties of the representations in such a way that following the rules preserves the contents of the representations. Hence, in answer to Q2, by taking mental processes to be computational, mental processes preserve content in just the way computational processes do, i.e., by obeying syntactic rules. The answer to Q1 then is that the contents of mental states derive from the contents of the mental representations underlying those states. In other words, the semantic properties of the objects of computation determine the semantic properties of the psychological states which are relations to those objects.

Of course, we have not *really* answered Q1 until we say how mental representations come to have content in the first place. This is a difficult and interesting question, but it is not a question which presently requires an answer. The present issue proceeds on the assumption that mental representations have contents and asks about the relation such contents bear to those of the relevant psychological states. If computational properties are not sufficient to determine intentional properties, then the putative answers to both Q1 and Q2 are simply unavailable.

Naturally, the last thing we ought to be willing to give up is the architecture. Were the Twin Earth intuitions sound, computational psychology would be a misnomer at best and a disaster at worst. The syntax of thought would become the proper domain of computer science and the semantics of thought would fall in the hands of philosophy with its thought experiments and counterfactual intuitions. The top two tiers of the architecture would then bear little, if any, interesting relation to each other, and the claim that computational psychology is a special science would thereby be without foundation. But the intuitions may not be sound. Twins may be more intentionally alike than many have supposed. If, as suggested below, this is so, then what is apparently required is some kind of revision or redescription of our pre-theoretical notion of intentional content, one which brings it more in line with the computational picture sketched above.

Let's take stock. The special sciences have a certain structure which, on the face of it, is shared by computational psychology. Part of this structure involves sufficiency relations between properties mentioned in the special sciences and properties of underlying mechanisms, broadly construed. But Twin Earth thought experiments are supposed to show that these sufficiency relations break down in the case of computational psychology. The reason is that identical physical and computational mechanisms can apparently have states with different intentional contents. This means that the contents which computational theories use in their explanations are not determined by other properties those theories mention; rather, they are determined by extra-theoretical considerations. So, according to this line of reasoning, computational psychology leaves its intentional properties unexplained. Thus, the problem of intentionality is left unsolved.

III

A natural strategy for resolving this difficulty suggests itself: show that, in an actual computational theory of some psychological domain, the intentional contents mentioned by the theory are determined by that theory. Computational psychology, at least for the putative domain, will then be shown to be a genuine special science, and the problem of intentionality as it is illustrated by Twin Earth thought experiments will be solved for that domain.

In order to show this, we need a starting point. We need, that is, some notion of content used by computational theories. To that end, consider Robert Cummins' notion of an s-representational content (1989, *passim*). Cummins distinguishes between s-representational content and intentional content. One reason for this is that s-representational contents are almost for free whereas full-blooded intentional contents are rather more expensive: "There is no 'further fact' required for successful s-representation beyond what is required for successful simulation" (1989, p. 99). So, for example, a representation *r* of *x* will s-represent *x*, the Gödel number of *x*, plus lots of other things, because there are systematic and coherent ways of interpreting *r* as *x*, as the Gödel number of *x*, and as lots of other things. Hence, failing to distinguish s-representational content from intentional content has crazy consequences like: my belief that France is in Europe can be a belief about a number. Moreover, the computational theory of cognition (CTC) Cummins develops has no internal resources for constraining s-representation in a way which yields the intuitively required intentional contents.

A further and more relevant reason Cummins makes the distinction is that the CTC is committed to *individualism* whereas intentional content is—according to the Twin Earth thought experiment mentioned above—*anti-individualist*. Anti-individualists typically hold that the contents of mental representations are determined "by reference to certain features of the world outside the subject, in such a way that the contents could not exist in environments that lacked those features" (Segal 1989, p. 190). Individualists typically deny this by insisting that contents do not vary across such environments, but that they are determined by other, environment-independent properties of the subject. Since the CTC is individualistic—it counts Twins as computationally identical—there is simply no way of arriving at differing intentional contents for Twins without imposing further constraints on s-representation *external to the CTC*. These further constraints typically embody relations to the environment in a way which prevents the CTC from employing genuine intentional properties in its explanatory apparatus.

There is, however, a way of finding the required constraints on s-representational content without reaching beyond some version or other of the CTC. The possibility of such a view is guaranteed by the following argument:

(i) Suppose we have support—*independently of the CTC*—for counting Twins as intentionally identical.
(ii) The problem of the relationship between the s-representational contents and the intentional content of a given symbolic representation is one of constraining the choice of the former to yield the latter.
(iii) In virtue of (i), the constraints mentioned in (ii) need not appeal to factors outside the CTC, since the constraints are only required to be external if (i) is false (this is why (i) must be independently established).

If this is correct, then we are free to investigate the CTC for the requisite constraints. In fact, I think this is correct, since the only potentially objectionable premise, (i), should not be objected to.[1]

Of course, we still do not know whether the CTC embodies strong enough constraints to make (i)–(iii) interesting. It could turn out that the CTC simply is not powerful enough. Nevertheless, I would like to suggest that in the case of one particular and limited CTC such constraints do exist.

The theory I have in mind is the computational theory of vision, as developed by Marr, Hildreth, Ullman, *et. al.*[2] Vision, according to this theory, is a process of inferences from relatively simple representations with relatively impoverished informational content to increasingly more complex representations with relatively sophisticated informational content. The visual system is therefore regarded as an information-processing machine. Such systems are to be understood at three levels. (1) The computational level "in which the performance of the device is characterized as a mapping from one kind of information to another, the abstract properties of this mapping are defined precisely, and its appropriateness and adequacy for the task at hand are demonstrated" (Marr 1982, p. 24). (2) The representational and algorithmic level which seeks a choice of representation and algorithm which effectively implements the information and processes specified by the computational theory. (3) The level of physical realization which details the way in which the representations and algorithms are physically realized. I shall be concerned only with the informational and representational parts of the overall theory.

At any given isolable stage in the computational process, the contents of the representations occurring at that stage are more or less tightly constrained by what is required for subsequent stages.[3] Marr writes, "stated baldly, the strong constraints come from what the representation is to be used for" (1982, p. 326). So what subsequent computations require in order to process information ultimately leading to a representation necessary for object recognition constrains the information represented in lower level representations. To take a simple example, a Marrian *edge* representation initially appearing in the raw primal sketch has an extension which includes such disparate physical phenomena as marks on a surface, thin shadows, creases, object boundaries, etc.[4] One reason for attributing a content with that extension is that subsequent grouping processes acting on *edge* representations use them to refer to increasingly abstract properties. If the assigned content is too fine-grained in the sense of being compatible with a narrower range of distal phenomena or differs in

---

[1] See Crane (1991), Segal (1989, 1991), and Seymour (1993) for a defense of the claim that Twin's are intentionally alike.

[2] I shall refer to it as Marr's theory, partly for ease of formulation and partly because the examples I use are from Marr's (1982) book, though of course quite substantial contributions have been made by many others.

[3] This idea is put to good use by Segal (1989) who argues that Marr's theory is individualistic.

[4] Marr is not clear whether the primitives in the raw primal sketch refer to properties of the viewed surface or to properties of the image on the retina. He sometimes suggests the latter (1982, pp. 41, 71, 91, 93, 366) and sometimes the former (1982, pp. 52, 68, 91). For ease of formulation, I shall adopt the former interpretation; but for present purposes, either choice will do. This is because on either interpretation, the raw primal sketch contains representations which remain intentional.

various respects from the actual content, the grouping processes simply will not be reliable.

The point is vitally important, since just these sorts of "top-down" constraints used at virtually every stage in the inferential process are going to nail down an s-representational content which has the correct semantics. The inference embodied in the grouping processes mentioned above is from the raw primal sketch to the full primal sketch. The raw primal sketch consists of primitive representational tokens, such as edges, blobs, and bars, derived by means of the spatial coincidence assumption from the zero-crossings obtained from the light intensity values on the retina. The full primal sketch is then derived by operating on the raw primal sketch with processes, such as selection and grouping, designed to make explicit spatial information which includes density, collinearity, and local parallelism. Grouping processes function to combine roughly similar types of representational primitives into larger primitives at greater scales designed to capture increasingly abstract properties of the viewed object. The process recursively builds upon itself depending on the complexity of structure the image possesses.

At a general level, it is already possible to see how such "top-down" constraints can rule out many of the possible s-representational contents of the primitives occurring in the full primal sketch. Suppose *edge* representations represented the Gödel number of surface markings, a number assigned to markings with certain properties. The grouping processes, then, would not be able to produce representations carrying information about collinearity or local parallelism or even spatial location since numbers simply lack these properties. Numbers are equal or inequal, greater or less than, added to or subtracted from, but they are not parallel or collinear. At a less general and more interesting level, it is also possible to see why *edge* representations occurring as inputs to the grouping processes have the contents they do, viz., as referring to such disparate physical phenomena as surface marks, creases, thin shadows, cracks, object boundaries, etc. Suppose, for example, that *edge* referred only to object boundaries. Then, assuming the spatial coincidence assumption (which can scarcely be coherently denied), the grouping processes would be regularly discerning objects where there were none. To use Marr's own example (1982, p. 91), an image of a striped cat viewed from, say, one foot would produce, after grouping, a representation of a distinct object for each hair (which is normal), a distinct object for each stripe (which is not), and a distinct object for the striped pattern (which is clearly absurd). So not only does the computational theory rule out "wild" contents like Gödel numbers, it also requires that tame contents, like object boundaries, cannot be the only ones.

The intention here is to show that it is possible to find tight enough constraints from within an actual computational theory to nail down an s-representational content from the available choices which has the correct semantics. Much more can be said about the case of vision. The later processes, such as the derivation of the $2^1/_2$-D sketch and the 3-D model representation, also make the point clear, though these processes are typically more complex. Stereopsis, for example, which forms part of the construction of the $2^1/_2$-D sketch, matches elements of the images from each eye to form a single image which contains disparity information used to determine the relative distances of objects from the viewer. In order to satisfy the constraints on available matches, object boundaries must form only a small part of the area of an image. Were this not so, disparity values would not remain constant enough to prevent numerous false matches. The process simply would not yield a reliable measurement of relative distance.

One might object to this line of reasoning on the grounds that the question of which contents the earlier representations possess is begged by simply assigning intuitively correct contents to later representations.[5] For example, the larger scale primitives produced by the grouping processes are intended to carry information about collinearity and local parallelism. But such informational content is more or less simply assigned. Why not assign a different content, such as the Gödel number of collinear elements? A computational theory of complex Gödel numbers could then be devised so that the constraints placed on the contents of earlier representations are no longer sufficient to rule out Gödel number contents.

There is, however, a clear and motivated answer to this objection. Marr's theory is revolutionary in many respects. One such respect concerns the proper method for understanding information-processing systems which requires a clear formulation of the purpose of the system under scrutiny. Essential to this formulation is a precise statement of the computational problem to be solved. In the case of vision, the overall problem is: given the representation of light intensity values on the retina as input, how does the system construct the complex descriptions of objects—3-D model representations—which are required for object recognition? It is this aspect of Marr's approach which separates his theory from those of earlier researchers (1982, *passim*, especially chapters 1 and 7). This

---

[5] The objection is due to Jerry Fodor, in conversation.

formulation of the problem is intentional, in that it refers to representations, descriptions, recognition, etc., and places strong constraints on the nature of the informational content of the output representation. Since the "top-down" constraints I have mentioned operate on the highest (output) representation, the question of the nature of the contents of the higher representations is not begged in the way the objection claims. Rather, the question is answered more or less pre-theoretically by the intentional definition of vision given by Marr.

Furthermore, the definition is not arbitrary. In chapter 1, Marr explains that it is inspired and supported by empirical and behavioral evidence presented by Elizabeth Warrington. She classified two sorts of patients with parietal lesions: those with lesions on the right side and those with lesions of the left. The former are able to recognize and identify a given common object from a conventional view, but if the view is unconventional—a clarinet seen end-on, for example—they would not be able to identify it and would even deny that it is the same object they just recognized from a conventional view. Patients of the latter sort, on the other hand, could often identify the shapes of objects even if the view was quite unconventional. This suggested to Marr, among other things, that vision alone produces representations of shape, space, and spatial arrangement. The definition, then, simply rules out representations of numbers as outputs.

IV

The computational theory of vision, then, is a special scientific theory in the same straightforward way as theories in geology, biology, and astronomy. All of them share the same sort of architecture, which is itself an interesting fact. And even though the theory employs intentional properties which are not associated with the other special sciences, it is rigorous, powerful, and probably true, at least in outline. In the course of sketching how this is all possible, we have seen why the visual system instantiates the intentional properties it does, and have thereby suggested a solution to the problem of intentionality.

Before these considerations are taken to be a complete solution to the problem of intentionality in all its guises, much work needs to be done searching for the analogous "top-down" constraints and intentional definitions in the case of mental states which do not operate more or less automatically, e.g., states like believing that gophers ruin golf courses. I do not pretend for a moment that this is an easy task. Nor do I pretend for a moment to undertake it here. So my conclusion is programmatic: thus far, we have no reason to think that these constraints do not exist—indeed, we have reason to think that they *do* exist, at least to the extent that if one intentional system embodies them, others probably do as well. So we should keep an eye out for them wherever we find intentionality.

Perhaps the most important aspect of the program also remains the same: the place to look is in actual computational theories of the domain in question.

Acknowledgments

Many thanks to Jerry Fodor, Deborah Mandelbaum, and Stephen Schiffer for their careful comments and criticisms of the ideas expressed in this paper. A special thanks to Gabe Segal to whom I am greatly indebted for his depth of insight and willingness to discuss his views with me.

References

Burge, Tyler 1979: Individualism and the Mental. *Midwest Studies in Philosophy*, vol. 4, pp. 73–121.

Crane, Tim 1991: All the Difference in the World. *Philosophical Quarterly*, vol. 41, no. 162, pp. 1–25.

Cummins, Robert 1989: *Meaning and Mental Representation*. Cambridge, Mass.: The MIT Press.

Fodor, Jerry 1987: *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, Mass.: The MIT Press.

Marr, David 1982: *Vision*. New York: W. H. Freeman and Company.

Putnam, Hilary 1975: The Meaning of "Meaning", in his *Mind, Language and Reality*, pp. 215–271. Cambridge, U.K.: Cambridge University Press.

Segal, Gabriel 1989: Seeing What Is Not There. *Philosophical Review*, vol. 98, no. 2, pp. 189–214.

Segal, Gabriel 1991: Defence of a Reasonable Individualism. *Mind*, vol. 100, no. 4, pp. 485–494.

Seymour, Daniel 1993: Some of the Difference in the World: Crane on Intentional Causation. *Philosophical Quarterly*, vol. 43, no. 170, pp. 83–89.