"Unexploited Content,"
in Graham McDonald and David Papineau, eds. Teleosemantics. Oxford University Press, 2006.

Jim Blackmon
David Byrd
Robert Cummins
Alexa Lee
Chris May
Martin Roth

I. Introduction

In this paper, we introduce a novel difficulty for teleosemantics, viz., its inability to account for what we call unexploited content—content a representation has, but which the system that harbors it is currently unable to exploit. In section two, we give a characterization of teleosemantics. Since our critique does not depend on any special details that distinguish the variations in the literature, the characterization is broad, brief and abstract. In section three, we explain what we mean by unexploited content, and argue that any theory of content adequate to ground representationalist theories in cognitive science must allow for it.[1] In section four, we show that teleosemantic theories of the sort we identify in section two cannot accommodate unexploited content, and are therefore unacceptable if intended as attempts to ground representationalist cognitive science. Finally, in section five, we speculate that the existence and importance of unexploited content has likely been obscured by a failure to distinguish representation from indication, and by a tendency to think of representation as reference.

II. Teleosemantics

Teleological accounts of representational content identify the extension of a representation R with the class of things C such that, historically, it was applications of tokens of R to members of C that led to the selection and replication of the mechanisms that produced or consumed tokens of R. Accounts along these general lines are familiar from the writings of Millikan, Neander, Papineau and others (Millikan 1984,

[1] There are, of course, initiatives in cognitive science that are not representationalist—e.g., the dynamic systems approach advocated by Van Gelder (1995) and others. If non-representationalist approaches ultimately carry the day, then disputes about how mental representation should be understood in cognitive theory will have been idle. For the most part, we simply assume in what follows that some form of representationalism is correct. But, now and again, we phrase matters more methodologically, as points about what representationalist explanations of cognition need to assume rather than as points about what cognitive processes actually require.

1986; Neander 1991; Papineau 1984, 1987; recent anthologies by Allen, Bekoff, & Lauder 1998; Buller 1999; Ariew, Cummins, & Perlman 2002). For our purposes, the crucial point about all such theories, is that a representation R can have the content C for a system only if it had, when selection took place, the ability to apply R to members of C. There cannot be selection for an ability that isn't there. The scenario underlying teleological accounts of content features a sub-population with a mechanism (what Cummins (1996a) calls an intender) in the R-application business. It applies R to a variety of things, including, under certain circumstances, members of C. Those applications – the applications of R to members of C – prove adaptive enough to cause the mechanism in question to spread through the population over time.

It is no part of this story that the reliability of R applications (applications of R to Cs vs. non-Cs) or their accuracy (excellence of fit between a token of R and its target C) is ever very good, or improves over time. All that is required is that it is the applications of R to members of C that leads to the spread of the mechanism through the population. The trait – the characteristic pattern of R applications – may go to fixation even though R is applied somewhat inaccurately to members of C, and only under rather special or rare circumstances, and frequently applied to other things. We emphasize this to make it clear that the critique we elaborate in the next section does not depend on the reliability or accuracy of the selected representing or representation consuming mechanisms. On the other hand, accurate enough applications of R to Cs cannot simply be random accidents: there must be a mechanism in the R-application business to select.

III. Unexploited Content

By unexploited content we mean information or content carried by or present in a representation that its harboring system is, for one reason or another, unable to use or exploit. A commonsense example will help to introduce the basic idea. Imagine someone who learns to use road maps to find a route from point A to point B. A study of the map might lead to the following plan: make a left at the third intersection, then another left at the next cross street, followed by an immediate right. It never occurs to this person to use the map to extract distance information until, one day, someone suggests that the map shows a shorter route than the one generated. Prior to this insight, our imaginary subject uses the map in a way that would be insensitive to various geometrical distortions, such as shrinking the north-south axis relative to the east-west axis. If assignments of representational content are limited by the abilities its user actually has to exploit the map, we will have to say that there is no distance information there to be exploited until after the user has learned to exploit it. And this will evidently make it impossible to explain how the user could learn to effectively compare routes for length: you cannot learn to exploit content that isn't there. Indeed, it is evident that this story makes no sense at all unless we concede that relative distances are represented before the user learns to exploit that information. Even if the user never exploits relative distance information, we are forced to allow that it is there to be exploited, since, under the right conditions, the user could have learned to use maps to compare distances. This would not be possible if the map did not represent relative distances.

How seriously should we take this sort of example? We think the lesson is far-reaching and fundamental. To begin with, the idea that a brain can learn to exploit previously unexploited structure in its representations is presupposed by all neural network models of learning. Such learning typically consists in adjusting synaptic weights so as to respond properly to input activation patterns. This whole process makes no sense unless it is assumed that input patterns represent proximal stimuli prior to learning, and that the representational content of input patterns remains the same throughout learning. Prior to learning, the network cannot properly exploit input representations: That is precisely what the process of weight adjustment achieves over time.[2]

Having come this far, we can see that the problem of learning to exploit "lower level" ("upstream") representations must be ubiquitous in the brain, if we assume that the brain acquires new knowledge and abilities via synaptic weight adjustment. In perceptual learning, for example, proximal stimuli must be represented before the appropriate cortical structures learn/evolve to exploit those representations in target location and recognition. As an example, consider the capacity to exploit texture gradients as visual depth cues. Representations in V1 contain texture gradients but the ability to exploit these as depth cues (as you do when you view figure 1) develops later (citation). Similarly, the ability to exploit retinal disparity in binocular vision as a depth cue develops along with the organization of binocular columns in the visual cortex. This process can be aborted by exotropy, but in such cases, binocular fusion

[2] We have heard it said that the network creates the content of its input patterns as learning progresses. But if we say this, we have no reason to say that early responses are errors. And if early responses are not errors, why change the weights in any particular direction? Indeed, why change them at all?

without stereoscopic depth vision can still be achieved as the result of surgical correction and vision training, demonstrating that the retinal disparity information is still present in early visual representations, but

figure one

There is no need to multiply examples. Once our attention is drawn to the phenomenon, it is clear that there must be many features of representations, especially structural features, at nearly all levels of perceptual and cognitive processing, that require learning and/or development for proper exploitation.

IV. Teleosemantics and Unexploited Content

Situating these facts in an evolutionary context immediately reveals a problem for teleosemantics. It is certainly possible, and probably common, that the abilities required to exploit various features of representations evolved well after those features appeared in the representations themselves. As just remarked, the ability to exploit texture gradients in early visual representation as depth cues might well have evolved well after well-defined gradients were available in those early representations. Now here is the point: The presence of texture gradients in early visual representations could not have been adaptive prior to the evolution of the processes that exploit them. Teleosemantics, however, implies that texture gradients did not represent depth until <u>after</u> it became adaptive for visual representations to include them. In general, content only becomes adaptive, hence a candidate for the kind of content-fixing selection contemplated in teleosemantics, when and if the ability to exploit it is acquired. Evidently, there can be no selection for an ability to exploit content that isn't there. The "opportunity" to evolve the ability to exploit texture gradients in visual representations as depth cues simply cannot arise unless and until depth-representing texture gradients become available to exploit.

Reflection on this last point suggests that the same difficulty arises whether the ability to exploit some feature of a representation is learned or evolved. For concreteness, assume that the ability to exploit texture gradients as depth cues is learned. While the current state of neuroscience provides no definitive account of such learning, it is perfectly intelligible to suppose it involves the systematic adjustment of synaptic weights in some sub-structure of the visual cortex. Evidently, if the ability to learn to exploit texture gradients itself evolved after texture gradients became available in early visual representations, we have a situation exactly like the one just rehearsed: Teleosemantics assigns no content to unexploited features of representations, and this undermines the obvious explanation of how the ability to learn to exploit such features might later become adaptive.

To sum up: Once our attention is directed to the phenomenon of unexploited content, it is natural to ask how the ability to exploit previously unexploited content might be acquired. Learning in the individual, and evolution in the species, are the obvious answers. Equally obvious, however, is that teleosemantics cannot allow for evolving the ability to exploit previously unexploited content: That requires content to predate selection, and teleosemantics requires selection to predate content.

V. Representation and Indication [3]

It seems likely that the very possibility of unexploited content has been overlooked in philosophical theories of content because of a failure to distinguish representation from indication. In this section, we digress a bit to explain how we understand this distinction, and conclude by suggesting how exclusive attention to indication tends to make the phenomenon of unexploited content difficult to discern.

1. Terminology. Some authors (e.g., Schiffer, 1987) use 'mental representation' to mean any mental state or process that has a semantic content. On this usage, a belief that the Normans invaded England in 1066 counts as a mental representation, as does the desire to be rich. This is not how we use the term. As we use the term, a mental representation is an element in a scheme of semantically individuated

[3] This section draws heavily from Cummins and Poirier (forthcoming).

types whose tokens are manipulated––structurally transformed--by (perhaps computational) mental processes. Such a scheme might be language-like, as the Language of Thought hypothesis asserts (Fodor, 1975), or it might consist of (activation) vectors in a multidimensional vector space as connectionists suppose (e.g., Churchland, 1995). Or it might be something quite different: a system of holograms, or images, for example.[4] An indicator, on the other hand, simply produces structurally arbitrary outputs that signal the presence or magnitude of some property in its "receptive field".

2. Indication. We begin with some influential examples.

- Thermostats typically contain a bimetallic element whose shape indicates the ambient temperature.
- Edge detector cells were discovered by David Hubel and Torsten Wiesel (1962). They write: "The most effective stimulus configurations, dictated by the spatial arrangements of excitatory and inhibitory regions, were long narrow rectangles of light (slits), straight-line borders between areas different brightness (edges), and dark rectangular bars against a light background."
- "Idiot lights" in your car have come on when, e.g., the fuel level is low, or the oil pressure is low, or the engine coolant is too hot.

'Indication' is just a semantic-sounding word for detection. Since we need a way to mark the distinction between the mechanism that does the detection, and the state or process that is the signal that the target has been detected, we will say that the cells studied by Hubel and Weisel are indicators, and that the pattern of electrical spikes they emit when they fire are indicator signals. Similarly the bimatalic element found in most thermostats is an indicator, and its shape is the signal.

3. Indication vs. representation. Indication is generally regarded as a species of representation. Indeed, causal and informational theories of representational content assert that representation is, or is inherited from, indicator content.[5] We think the two should be kept distinct.

[4] It is possible that the brain employs several such schemes. See Cummins (1996b), and Cummins et. al (2001) for further discussion of this possibility. [5] The theory is generally credited to Denis Stampe (1977). Its most prominent advocates are Fodor (1987) and Dretske (1981).

Indication is transitive, representation is not. If S3 indicates S2, and S2 indicates S1, then S3 indicates S1. Imagine a photo-sensitive cell pointed at an "idiot light" in your car, and attached to a relay activating an audio device that plays a recording: "The oil pressure is low." If the light indicates low oil pressure, so does the recording. Representation, on the other hand, is not transitive. A representation of the pixel structure of a digitized picture of the Statue of Liberty is not a representation of the statue's visual appearance, though the later may be recovered from the former.[6] To anticipate some terminology we will use later, a representation of the pixel structure is an encoding of the statue's visual appearance.[7]

Indicator signals are arbitrary; representations are not. This is implied by the transitivity of indication. Given transitivity, anything can be made to indicate anything else (if it can be detected at all), given enough ingenuity and resources. Because indicator signals are arbitrary, disciplined structural transformations of them cannot systematically alter their meanings. Such transformations, however, are precisely what make representations useful. Consider, for example, a software package that takes a digitized image of a face as input and "ages" it, i.e., returns an image of that face as it is likely to look after some specified lapse of time. Nothing like this could possibly work on an input that was required only to indicate a certain face -- a name, say -- because there is no correlation between the physical characteristics something must have to be a signal that indicates the appearance of my face at age 18 and the physical characteristics of my face at age 18. It follows from the nature of indication that the structural properties of an indicator signal have no significance. Indicators "say" that their targets are there, but do not "say" anything about what they are like. Representations, on the other hand, mirror the structure of their targets (when they are accurate), and thus their consumers can cognitively process the structure of the target

[6] Representation, on the view advocated by Cummins (1996a), is grounded in isomorphism. Since isomorphism is plainly transitive, it might seem that representation must be transitive too. In a sense, this is right: the things that stand in the isomorphism relation are structures -- sets of "objects" and relations on them. If S1 is isomorphic to S2, and S2 is isomorphic to S3, then S1 is isomorphic to S3. An actual physical representation, however, is not an abstract object; it has a structure -- actually, several -- but it isn't itself a structure. A connected graph structure of a paper road map is isomorphic to the street and intersection structure of a town, but not to the town's topology. The town's topology is isomorphic to the topology of a citrus grove. But no structure of the road map need be isomorphic to any structure of the grove. [7] It is what Haugeland would call a recording of the picture. See Haugeland, 1990)

by manipulating the structure of its representation. But representations, unlike indicator signals, are typically silent concerning whether their targets are "present": they are not, except incidentally and coincidentally, detector signals.

<u>Indicators are source dependent in a way that representations are not</u>. The cells studied by Hubel and Weisel all generate the same signal when they detect a target. You cannot tell, by looking at the signal itself (the spike train), what has been detected. You have to know which cells generated the signal. This follows from the arbitrariness of indicator signals, and is therefore a general feature of indication: the meaning is all in who shouts, not in what is shouted.[8]

In sum, then, indication is transitive, representation is not. It follows from the transitivity of indication that indicator signals are arbitrary and source dependent in a way in which representations are not, and this disqualifies indicator signals as vehicles for structure dependent cognitive processing. Representation is intransitive, non-arbitrary and portable (not source dependent), and therefore suitable for structural processing. Indicator signals "say" their targets are present, but "say" nothing about them; representations provide structural information about their targets, but do not indicate their presence. Indicator signals say, "My target is here," while representations say, "My target, wherever it is, is structured like so."

4. Discussion. If indication is your paradigm of mental content, as it is bound to be if you hold some form of causal theory, you are going to focus on what fixes the content of an indicator signal.[9] Whatever fixes the content of an indicator signal, it is not its structural properties. In this context, therefore, motivation is lacking for thinking about which aspects of a representation's structure can usefully be processed, and whether the ability to do that processing is learned or evolved or a combination of both. Maps rub your nose in the possibility of unexploited content; idiot lights do not.

There can, however, be unexploited indicator signals. Think of the color coded idiot lights at intersections: you have to learn that red means stop, green means go. This is also unexploited content (though not what

[8] We do not mean to imply here that the shape of a spike train is never significant. The point is rather that two indicators can have the same spike train, yet indicate different things. [9]See Cummins (1997) for more on the marriage between causal theories, indication and the language of thought.

we have been calling representational content), and, unsurprisingly, it makes trouble for teleosemantics. Teleosemantics implies that an indicator signal has no content until there has been selection for the indicator that generates it. But the ability to exploit, or to learn to exploit, an indicator signal can only evolve if the indicator is already there signaling its target.

Magnetosomes (an example featured in Dretske 1986) are magnetically polarized structures (typically ferrite surrounded by a membrane) in single-cell ocean dwelling anaerobic bacteria. The orientation of these structures correlates with the direction of the earth's magnetic field. By following the magnetic orientation in a particular direction, organisms far from the equator can avoid aerobic water near the surface. For this to work, magnetosomes must be chained and attached at both ends of the cell to form a reasonably straight line in the direction of locomotion (see figure 1). This is because the orientation of the organism is simply a consequence of the orientation of the chain of polarized molecules. The whole body of the bacterium is a floating compass needle. The organism swims, and will move in whatever direction

Figure 2: Magnetotactic bacterium from the Chiemsee, Bavaria,Germany (Biomagnetism Group, University of Munich). Dark blobsare sulfur granules.

Chaining, of course, is simply a physical consequence of having a lot of little magnets suspended in proximity. They will stick together north to south. What is not so obvious is why the north pole of the string winds up attached at the "front"—i.e., direction of locomotion—end of the organism. However this happens, it is certainly possible, indeed probable, that the attachment process evolved after magnetosomes themselves appeared within the cell body of anaerobic bacteria. Selectionist theories imply that magnetosome chains did not indicate the direction of anaerobic water until after it became adaptive to do so, i.e., only after the evolution of attachment. But surely it is in part because they did indicate the direction of anaerobic water, that the attachment process was adaptive enough to be selected for.

VI. Conclusion

A very natural response to the forgoing is to say that unexploited content isn't really content. After all, there is a lot of unexploited information in the environment, information that cognitive systems must acquire the abilities to exploit. We do not call that information content.

We are sympathetic with the comparison between learning/evolving an ability to exploit information in a representation or indicator signal and learning/evolving an ability to exploit information in the environment. We think these are, in fact, deeply similar. The importance of this similarity is obscured or lost in theories that essentially take representation to be reference. Theories of content that take representation to be reference perforce focus on devising the conditions that (allegedly) fix the references of semantically primitive terms, relying on the standard truth-conditional combinatorics to fix the references and truth-conditions of complex expressions. Access to something that refers to horses—a primitive term in Mentalese—however, tells you nothing about horses. Actual information about horses, therefore, is to be found only in the (or a) set of Mentalese sentences that contain a |horse| (a Mentalese term referring to horses) and appear in the Belief Box. The only sense in which such an account allows for unexploited content, therefore, is the sense in which a cognitive agent might not exploit all of its beliefs about horses on a particular occasion. While this is undoubtedly a case of unexploited information, it is not a case of the sort we have been discussing. Returning to our analogy, inability to extract relative distance information from a road map is very different than failing to read or utilize some of the sentences in a book. In the later case, the content of the unread sentences is unproblematically extractable from those sentences; they just are not noticed for one reason or another. The problem is not that one doesn't know how to read them. In the case of the map, a new skill is required to exploit the needed information. Unexploited information of the sort allowed for in Language of Thought theories evidently poses no problem for teleosemantics comparable to the one we have been urging, since the mechanisms responsible for applying the primitive constituents and processing the relevant syntactical machinery may be selected for independently of their occurrence in any particular belief.

Cognitive systems need information. LOT accounts attempt to provide for this by giving indication a two-fold role. First, indicator signals alert the organism to the instantiation of their target properties in their receptive fields. [10] Second, primitive terms of LOT inherit their references from the properties they are used to indicate in the context of detection. Cognition is understood as the application of theories expressed as organized sets of sentences in Mentalese, hence as a species of truth-conditional inference, implemented as computation over symbolic structures with Tarskian logical forms.[11]

Perhaps something like this story makes sense for the "higher" cognition of adult humans, especially if, like Plato, one is inclined to suppose that cognition is thinking and that thinking is essentially talking to oneself. But this picture is of little use for understanding phenomena like the capacity to exploit texture gradients in early visual representations as depth cues. When we turn to phenomena such as these, the truth-conditional semantics of propositional attitude contents is of dubious significance to cognitive science. Much more important, we believe, is the conception of representation and indication briefly introduced above. Representations, thus conceived, are of use to cognitive systems as an information source in addition to the environment (i) because they can be stored, and (ii) because they can be structurally transformed in ways that the environment typically cannot be. Sophisticated indicators are of use because they can signal the presence of environmental features—e.g., the presence of a predator or a mate—that are extremely abstract from the point of view of proximal stimulation. Representation and indication thus conceived are what make reference and the propositional attitudes possible. A theory that begins with the truth-conditional semantics of propositional attitude contents thus skips over most of the action and begins with a phenomenon that is just the sort of cognitive achievement that mainstream cognitive science seeks to explain in terms of representation.

[10] See Cummins and Poirier (forthcoming) for a discussion of how indicators might be come "source-free" and function as terms. [11] A paradigm example is Pollock (19@@).

We do not wish to quibble over whether the phenomenon we have called unexploited content is really content. We do contend, however, that what we are calling content is what ultimately does the work in representationalist cognitive science. No doubt we need to mark the distinction between exploited and unexploited content. We think 'exploited content' and 'unexploited content' do the job nicely. Refusing to use the word 'content' for as yet unexploited features of structured representations strongly suggests, wrongly, that those features are somehow different than those that are exploited. There is no intrinsic difference between texture gradients that are exploited and texture gradients that are not. To suppose otherwise would be like supposing that road maps cease to represent relative distances in the hands of those who cannot extract that information from them.[12][13]

In this paper, we have urged what we think is a novel objection to teleosemantic theories, namely that they cannot accommodate unexploited content/information. Surely, a necessary condition for the plausibility of a theory of mental representation that hopes to ground representationalist cognitive science is that it accommodate unexploited content/information. For it must be possible for a system to be able to learn or evolve the capacity to exploit the information carried by a representation or indicator signal, and this implies that the information is there prior to acquisition of the capacity to exploit it.

Works Cited

Allen, C., Bekoff, M., & Lauder, G. (eds.) (1998) *Nature's Purposes,* MIT Press.

Ariew, A., Cummins, R., & Perlman, M. (eds.) (2002) *Functions: New Essays in the Philosophy of Psychology and Biology*, Oxford UP.

Buller, D. (ed.) (1999) *Function, Selection, and Design*. SUNY Press.

Churchland, P. (1995), *The Engine of Reason, the Seat of the Soul*, MIT Press.

[12] There is a temptation to think that an unexploited feature of a representation doesn't represent anything *to* (or *for*) the system that harbors it. This is surely right. But to assimilate representation to *representation to/for* will, like teleosemantics, make it impossible to understand how, e.g., the ability to exploit texture gradients as depth cures could be learned or evolved. For more on the representation/representation-to distinction, see Cummins (1996a) [13] Notice, by the way, that what is important about texture gradients is not just that they somehow covary with depth. It is their suitability for structural processing that makes them useful. When covariation is all that matters, an arbitrary indicator signal is all that is required.

Cummins, R. (1996a), *Representations, Targets, and Attitudes*, MIT Press.

Cummins, R. (1996b) “Systematicity,” Journal of Philosophy, 93: 591
614.

Cummins, R. (1997) "The LOT of the Causal Theory of Reference." The Journal of Philosophy, 94: 535-542.

Cummins, R., J. Blackmon, D. Byrd, P. Poirier, M. Roth, and G. Schwarz (2001), “Systematicity and the Cognition of Structured Domains,” Journal of Philosophy, 98:167-185.

Cummins, R., and Poirier, P. (forthcoming), “Representation and Indication” in Representation in Mind, ed. Staines, P., Elsevier.

Dretske, F. (1981), Knowledge and the Flow of Information, MIT Press.

Fodor, J. (1975), The Language of Thought, Thomas Y. Crowell.

Fodor, J. (1987), Psychosemantics: The Problem of Meaning in the Philosophy of Mind, MIT Press.

Haugeland, J. (1990), “Representational Genera,” in W. Ramsey, S. Stich, and D. Rumelhart, (eds.) Philosophy and Connectionist Theory, Lawrence Erlbaum Associates.

Hubel, D. and Wiesel, T. (1962), “Receptive fields, binocular interaction and functional architecture in the cat's visual cortex.” Journal of Physiology of London, 160:106-154.

Millikan, R. (1984), Language, Thought, and Other Biological Categories, MIT Press.

Millikan, R. (1986), “Thoughts Without Laws: Cognitive Science Without Content,” Philosophical Review 95: 47-80.

Neander, K. (1991), “The Teleological Notion of Function,” Australasian Journal of Philosophy, 69, 454-468.

Papineau, D. (1984), "Reality and Explanation," Philosophy of Science 51: 550-572.

Papineau, D. (1987), Reality and Representation, Basil Blackwell.

Schiffer, S. (1987), Remnants of Meaning, MIT Press.

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.

Van Gelder, T. J., & Port, R. (1995) "It's About Time: An Overview of the Dynamical Approach to Cognition," in R. Port & T. van Gelder ed., Mind as Motion: Explorations in the Dynamics of Cognition, MIT Press, 1-43.