

MODERATE HOLISM AND THE INSTABILITY THESIS*

Mental or semantic *holism* can be roughly characterized as the doctrine that the meanings of one's sentence or the contents of one's beliefs are a function of (are determined by) their relations to all of one's other sentences and beliefs. Holistic theories of meaning and content are typically criticized for (among other things) leaving the ideas of translation, disagreement and change of mind problematic. However, such criticisms are more properly directed at an "instability thesis" which, while often taken to be a consequence of holism, can be separated from it. A 'moderate' holist need not be committed to the instability thesis, and can thus avoid many of the problems traditionally associated with holism.

Holism has two alternatives: atomism (the doctrine that each of one's beliefs or sentences gets its content independently of its relation to one's other beliefs or sentences) and molecularism (the doctrine that each of one's beliefs or sentences gets its content in virtue of its relation to a corresponding subset of one's other beliefs or sentences). Atomistic and molecular theories face a number of well-known problems.¹ Nevertheless, they are often favored over holistic theories because holism seems to demand that *any* differences in belief must result in differences in meaning or content. Among the most frequently cited reasons for not being a holist are the following two consequences of this purported demand.²

First, holism seems to make literal translation from one person's language into another's impossible. No two people have exactly the same beliefs, so by insisting that what one means by one's terms is a function of all of one's beliefs, holism suggests that no two people mean quite the same thing by any of their terms. If what one meant by, say, "elephant" were determined by all of

* I'd like to thank Robert Brandom, Joe Camp, Johnathan Cohen, Joshua Finkler, John McDowell, Madeline Muntersbjorn, Ram Neta, Gerald Vision, the editor of this journal, and members of the audience at the 1998 APA Central Division Meeting for comments on earlier versions of this paper.

¹ Its commitment to the possibility of having, say, the concept *red* without the concept *color*, along with its difficulties in arriving at a satisfactory account of error, have made atomism seem implausible to many. Molecularists, on the other hand, have problems providing criteria to distinguish those relations among beliefs that are meaning-constitutive from those that are not (problems familiar since Quine's attack on the analytic/synthetic distinction), and thus have difficulty preventing any non-atomistic position from sliding into a full-blown holistic one. Whether such criticisms are indeed telling is an open and important question, but since it is not my purpose here to criticize these non-holistic views, I will not pursue it here.

² For a much more exhaustive list of purported problems with holistic theories (including their apparently disastrous consequences for psychological and scientific explanation), see Fodor and LePore 1992.

one's elephant-beliefs, then someone who believed that "elephants are afraid of mice" would mean something different by "elephant" than someone who disbelieved it. As a result, there would be a radical 'incommensurability' between people with different beliefs. Unless two people shared *all* their beliefs, they wouldn't share *any* of them. For instance, if Peter believes that elephants are afraid of mice, and Mary doesn't, then what may have initially appeared to be shared beliefs -- both claim "elephants are big animals" -- would turn out to have different contents in virtue of the two believers meaning different things by "elephant." Consequently, a translation from one of their languages into the other's couldn't be completely accurate, and their ascriptions of belief to each other would be, strictly speaking, false. Complete communication would thus be effectively impossible, and we would never fully grasp the content of anyone else's words or thoughts. Worse still, the same phenomenon occurs with even our most recently past selves. Since our total set of beliefs is constantly changing, the holist seems committed to saying that the meanings of all of our terms and the contents of our beliefs are constantly changing as well. Translations within a single person's idiolect from one time to another would thus be inaccurate, and our belief ascriptions to our own past selves would also be, at best, only roughly true.

Similarly, holism seems to make disagreement over matters of fact impossible. Peter and Mary might have *seemed* to disagree over a matter of fact about elephants, but the holist can only treat their 'disagreement' as being over whether or not to accept the sentence "elephants are afraid of mice." Since the two would mean different things by the sentence, their 'disagreement' would just be a case of their talking past one another. There is no proposition that one believed to be true and the other believed to be false. Holism also seems to make it impossible to disagree with one's past self, and would thus rule out the possibility of one's ever changing one's mind about a proposition. If I initially believed that 'elephants' were scared of mice, and subsequently came to believe that 'elephants' were not scared of mice, then what I meant by "elephant" would have changed. There would thus be no proposition that I believed at one time and came to disbelieve at another.

Such purported consequences of holism are extremely counterintuitive since we take translation, disagreement and change of mind to be relatively unproblematic. Furthermore, most people take it to be simply *obvious* that there can be differences of belief without differences of meaning. Consequently, the holist's purported denial of this possibility makes his view seem extremely unattractive.

Short of giving up holism in favor of atomism or molecularism, there are two popular responses to the problems raised above.³ One is to adopt a 'two factor' theory in which a 'narrow content' is determined holistically while a non-holistic 'wide content' is comparatively invariant and shared.⁴ Disagreement, translation, and change of mind would thus all be understood in terms of the type of content that is not holistic. The other response is to suggest that we replace talk of content identity and difference with talk of content similarity.⁵ A single belief change may affect the content of all the terms involved, but the *pre* and *post* change contents will remain, nevertheless, very similar.

Both of these responses, however, concede that the holist is committed to what will here be called the "instability thesis," namely:

Any change in one's attitude towards a sentence will change the meanings of the terms contained in it and the contents of the associated beliefs.

Since the instability thesis requires that there be no changes of belief without changes of meaning, its truth would leave meaning and content tremendously unstable, constantly changing even within a single person. The instability thesis lies at the heart of people's intuitive difficulties with holism, and a large part of the appeal of molecularism and atomism results from their not having to accept it. Nevertheless, holism's commitment to the instability thesis is assumed by its defenders as often as by its critics.⁶ Both of the responses mentioned above suggest ways for the holist to try to live

³ A third response, simply biting the bullet and accepting all the consequences come what may, is less popular and will not be discussed here.

⁴ See, for instance, Block 1986 and Field 1977.

⁵ See, for instance, Block 1986, Churchland 1986, and Harman 1973, 1993.

⁶ Instances of this assumption are too numerous to exhaustively list. Consequently, I will simply note that the connection between holism and instability is not only assumed by critics of holism, such as Fodor and Lepore (1992) and Devitt (1996), but is taken for granted in sympathetic 'expository' accounts of what holism is. For instance, the entries for holism in *The Routledge Encyclopedia of Philosophy* and *The Blackwell Companion to the Philosophy of Mind* (Block 1998 and Guttenplan 1994) both take it as uncontroversial that holism leads to instability. It is also noteworthy that critical responses to Fodor and Lepore's attack on holism never question

with the instability thesis. The first admits the instability thesis to be true of one type of content, and then looks for a second non-holistic type of content for which it doesn't hold.⁷ The second suggest that we accommodate the instability thesis by realizing that we can replace talk of *sameness* of meaning or content with talk of *similarity* of meaning or content, and recognize that the thesis doesn't leave meaning *that* unstable.

However, such responses have not been entirely successful in making the instability associated with holism palatable. Indeed, it has been questioned whether such positions are even coherent. In particular, some have challenged the intelligibility of the claim that there are independent wide and narrow contents,⁸ and others have argued that one cannot replace content identity with content similarity because the former concept is presupposed by the latter.⁹ Whether or not such criticisms are ultimately telling, it remains the case that the instability thesis *is* hard to live with, and it isn't clear that *any* attempt to make it palatable will be successful.

Fortunately, such attempts to make the instability thesis palatable may not be necessary because it isn't at all obvious that every holist must embrace instability. Holism only requires that the content of any one of one's beliefs *depend upon* or be a *function of* one's other beliefs, and *this* claim need not commit one to the instability thesis. After all, one can claim that *A* is a function of *B* without implying that any change in *B* will produce a change in *A*. Consider, for instance, the claim that one's final letter grade in a class is a function of (depends upon) the results of one's exams, quizzes and homeworks. The truth of this claim certainly doesn't entail that no two people could have the same final grade unless they had precisely the same score on all of their homeworks, exams and quizzes. Neither does it entail that any change to one of one's quiz scores will produce a change in one's final grade. Each result makes *some* contribution to one's final grade, but not every change among the contributors will produce a corresponding change in the

the link they draw between holism and instability (see the papers in Fodor and LePore 1993, and the symposium on their book in *Philosophy and Phenomenological Research* (Vol LIII, No. 3, Sept 1993)). For a recent debate about holism that takes instability for granted, see Lormand 1996, Becker 1998 and Margolis and Laurence 1998.

⁷ Similar moves are made (but without any commitment to a narrow/wide distinction) in Bilgrami 1992 and Lormand 1996.

⁸ See McDowell 1986, Bilgrami 1992.

⁹ See Fodor and LePore 1992, especially pp. 17-22.

ultimate outcome. The function from contributing scores to final grades is *many-to-one*, and thus allows a good deal of stability in the output in spite of the possibility of tremendous variation in the input.¹⁰

In much the same way, holism would entail instability *only* if it required that the function from beliefs to contents be *one-to-one*, and there is nothing in the general characterization of holism that requires that the type of function involved be restricted in this way. If the function from belief to meaning turns out to be many-to-one, then content could be comparatively stable through changes in belief. One could thus claim that the meaning of “elephant” is a function of one’s “elephant” beliefs without implying that every change in one’s “elephant” beliefs will produce a corresponding change in the meaning of “elephant.” Consequently, there is no reason why *simply* being a holist about meaning or content need commit one to the instability thesis.

We can thus distinguish *moderate* versions of holism, which treat the function from belief to meaning as many-to-one (and thus allow variations in belief that fail to produce variations in meaning), from *radical* versions of holism, which treat the function as one-to-one (and thus require any change in belief to produce a corresponding change in meaning). It is only radical holism that entails the instability thesis. A version of holism is thus best understood as either radical or moderate in virtue of its understanding of the nature of the function from beliefs to meanings, not in virtue of whether it places any restrictions on the inputs to this function.¹¹

Of course, many holistic theories of meaning and content may be of the more radical sort, and thus may be committed to the instability thesis. For instance, those functionalists who identify a thought’s content with its functional/inferential/conceptual role may (in the absence of criteria for dividing meaning constitutive inferences from others) be committed to a type of holism inseparable from the instability thesis.¹² In general, when holism is combined with a type of semantic

¹⁰ Of course beliefs and grades work in very different ways, and while grading is a good example of a many-to-one function from which some useful parallels can be drawn, one shouldn’t push the analogy too far.

¹¹ The latter approach would amount to equating moderate holism with a type of molecularism. For such a use of “moderate holism,” see Guttenplan 1994.

¹² See, for instance, Block 1986, Field 1977, Harman 1973, and Sellars 1974. A similar type of holistic theory that may entail a commitment to the instability thesis is Churchland’s State-Space Semantics. (See Churchland 1979, 1986.)

internalism, the instability thesis tends to follow. It is much less likely to do so when holism is combined with a type of semantic *externalism*. Our internal states are idiosyncratic and changeable, and so the internalist is well placed to understand meaning as the type of thing that the instability thesis could be true of. Instability can be accommodated easily within an internalist framework because when our beliefs change, our internal states typically change as well. The internalist is thus free to understand the ‘function’ from belief to meaning as the one-to-one function of *identity*: the meaning of a term just *is* the set of beliefs, attitudes and inferences associated with it. Externalists, on the other hand, typically understand meaning at least partially in terms of belief-independent and ‘public’ factors that are less likely to change with the agent’s beliefs. Consequently, the function from beliefs to meanings is more likely to be many-to-one and map beliefs onto a shared set of comparably stable external objects.

One can see an example of this in the ‘interpretational’ or ‘confirmation-based’ types of holism inspired by Quine’s remark that “the unit of empirical significance is the whole of science.”¹³ That such holistic accounts need not be committed to the instability thesis becomes clear when we consider Davidson’s suggestion that (very roughly) we identify the content of a sentence with its truth-conditions and the meanings of various terms with the (sets of) objects they refer to. While the suggestion that meaning be understood in terms of reference is frequently associated with atomistic theories, such an account can be holistic if reference (and hence meaning) is itself determined holistically.

One can arrive at such a holistic account of reference by following N. L. Wilson’s Principle of Charity, which suggests that one take the reference of a speaker’s terms to be whatever set of objects maximizes the total (or some weighted total) number of sentences the speaker truly believes.¹⁴ That is to say, the function takes the total set of sentences held true as its input and gives beliefs and concepts/referents as its output. Such an account of reference will undoubtedly be holistic: a term picks out the object it does because of the role that object plays in contributing,

¹³ Quine 1951, p.42. Many writers in this tradition (most notably Davidson and Quine) have, admittedly, flirted with the instability thesis, but in spite of these flirtations, their actual case for holism can be separated from it.

¹⁴ Wilson 1959.

either directly or indirectly, to the truth of countless beliefs.¹⁵ Nevertheless, even if countless beliefs played some role in determining the term's reference, there is little reason to think that a change in one (or even a considerable number) of these beliefs will change what is referred to. The truth of two different sets of 'elephant' beliefs may, in spite of their differences, be maximized by precisely the same set of objects. Such a holist about reference could thus allow there to be changes in belief without changes in meaning, and hence avoid any commitment to the instability thesis.

The Principle of Charity has, of course, been subject to fairly heavy criticism. In particular, it has frequently been criticized for misidentifying the truth conditions of our utterances. For instance, if I systematically identify salamanders as "lizards," the Principle of Charity might seem to dictate that the truth conditions for my sentence "this is a lizard" must include salamanders. While I will not be able to give a thoroughgoing defense the Principle of Charity here,¹⁶ I'll only mention that the principle will be comparatively resistant to counterexamples (and turn out to be compatible with most familiar forms of externalism) once one recognizes that:

- (1) The Principle is meant to "maximize" the true beliefs of the speaker in a way that is not strictly numerical (one thus needn't worry about the Quixotic project of counting beliefs). What is equally important to the goal of maximization is the comparative entrenchment of the beliefs that are held. Treating a central belief as false to preserve the truth of two peripheral beliefs is far from charitable.
- (2) Charity is a 'global' (or 'holistic') principle, and its local application is often misleading.
- (3) Some of the beliefs that the principle deals with, indeed some of the most deeply held ones, are rarely made explicit.¹⁷

¹⁵ Of course, certain beliefs will be (with respect to each word) more heavily weighted than others (just as exams are more heavily weighted than quizzes), but (*pace* the molecularist) one should not think that these heavily weighted elements are the only things that contribute to meaning. Only some beliefs may have an immediate effect on meaning, but enough changes to the comparatively unimportant beliefs can occasionally change what is meant as well.

¹⁶ I attempt to do so in Jackman 1999.

¹⁷ Indeed, some of these 'implicit' commitments may not even be best described as beliefs (see Jackman 1996, 1999).

Once Charity is understood this way, we can see why one need not worry about cases like the salamanders mentioned above. My belief that a given salamander is a lizard is not as central for me as my beliefs that all lizards are reptiles, or, say, my assumption that I mean the same thing by “lizard” as most zoologists. Treating my utterance of “This is a lizard” as true if and only if I’m looking at a salamander-or-a-lizard would make these more deeply held general beliefs false. Consequently, such a disjunctive interpretation of “lizard” would be uncharitable. Or, to take another case, while putting the condensation from various air conditioners into the extension of “rain” would make a few more of my “its raining” utterances true, it would be uncharitable because it would falsify much more deeply held beliefs of mine such as “rain doesn’t come from air conditioners.”¹⁸

The fact that the type of moderate holism outlined above makes use of the Principle of Charity is not, then, a serious problem for it, and none of the purported problems with holism discussed earlier face such a moderate holist. Two people can mean the same things by their words, in spite of differences in their beliefs, because the function from their beliefs to what they mean is one that yields the same output from their often significantly different inputs.¹⁹ While no two speakers may have precisely the same ‘elephant-beliefs,’ the objects that maximize the truth of their belief sets may be the same. As a result, both disagreement over matters of fact and translation from one idiolect to another become unproblematic. Much the same can be said for changes of mind. If I start believing that elephants are afraid of mice, I change one minor input into my ‘elephant-function,’ but not one that is important enough to change the function’s output.²⁰ The set of objects that maximizes the number of true ‘elephant’-beliefs that I have (presumably elephants) will not change just because I start believing that elephants are afraid of mice. Only rarely will changes in a few beliefs change the object or objects that maximizes the truth of the belief set. Meaning can

¹⁸ Such cases are, once again, discussed in considerably more detail in Jackman 1999.

¹⁹ That is to say, the fact that the function is many-to-one in no way requires that the inputs be ‘similar’ in the sense that Fodor takes to be problematic. The ‘conceptions’ people associate with a word could be radically different and still be satisfied by the same objects.

²⁰ Still, one should not assume that the lack of immediate effect on meaning produced by such belief changes entails that those beliefs didn’t contribute to what was meant. Such an assumption would be no more justified than assuming that the lack of immediate affect on one’s final grade produced by some quiz score changes entailed that those quizzes did not contribute to the final grade.

thus be invariant through belief change, and changes in attitude towards a single proposition will be possible

There remains, of course, the familiar problem of whether truth-conditional accounts of content, whether they are holistic or not, can account for, say, propositional attitude ascriptions. I obviously don't have space to deal with this important but distinct question here, especially since the main purpose of this paper is to argue that holists need not be committed to instability, not to defend a particular holistic account of content.²¹ Nevertheless, one should note that even if 'fine grained' conceptions must be added to the 'coarse grained' truth-conditional contents, the position outlined here (unlike typical 'two factors' accounts) understands *both* factors holistically.²² Consequently, working with both concepts and conceptions would not amount to a retreat from holism. Furthermore, once holistic truth-conditions are in place, claims about the similarity of the finer grained conceptions are relatively unproblematic. Fodor argues that holists cannot help themselves to the idea of similarity because similarity judgements presuppose identity relations between at least some of the beliefs that make up the conceptions.²³ However, while Fodor assumes that holists are not entitled to presuppose such identity relations, if one identifies beliefs by their holistically determined truth-conditional content, then one does have the resources to explain similarity relations between conceptions. Conceptions are similar to the extent that they are made up of beliefs that have the same truth-conditions.

In conclusion, then, many of the criticisms of holism that have lead people to embrace either atomism or molecularism are not criticisms of holism per say, but only of the instability thesis characteristic of certain radical brands of the doctrine. More moderate varieties of holism, which treat the function from belief to meaning as many-to-one (and thus don't require that every change in belief produce a change in meaning) are unaffected by such criticisms.

²¹ The sort of account suggested above is, however, defended in Jackman 1996.

²² Furthermore, given that the beliefs that make up the conceptions also determine the referents, one need not worry about "the nasty question: *What keeps the two notions of content stuck together?*" that Fodor and LePore claim cannot be answered by traditional 'two factor' accounts (Fodor and LePore 1992, p. 170). For a suggestion that traditional accounts such as Block's also have an answer to Fodor and LePore's challenge, see Bilgrami 1998.

²³ Fodor and LePore 1992, pp. 17-22.

References

- Becker, K. 1998. "On the Perfectly General Nature of Instability in Meaning Holism," *The Journal of Philosophy* Vol. XCV, n.12. pp. 635-640.
- Bilgrami, A. 1992. *Belief and Meaning*. Cambridge: Blackwell.
- Bilgrami, A. 1998. "Why Holism is Harmless and Necessary," in Tomberlin, (ed.), *Philosophical Perspectives 12: Language, Mind and Ontology*. Boston: Blackwell 1998.
- Block, N. 1986. "Advertisement for a Semantics for Psychology," in French, Uehling & Wettstein. 1986. *Midwest Studies in Philosophy X: Studies in the Philosophy of Mind*. Minneapolis, University of Minnesota Press.
- Block, N. 1998. "Holism, Mental and Semantic," in Craig, (ed.), *The Routledge Encyclopedia of Philosophy*. London: Routledge, 1998.
- Churchland, P. 1979. *Scientific Realism and the Plasticity of Mind*. Cambridge: CUP.
- Churchland, P. 1986. "Some reductive strategies in cognitive neurobiology," in Churchland, P. A *Neurocomputational Perspective*. Cambridge: MIT 1989.
- Devitt, M. 1996. *Coming to Our Senses*. New York: CUP.
- Field, H. 1977. "Logic, Meaning and Conceptual Role" *The Journal of Philosophy* Vol. LXXIV, n.7, pp.379-409.
- Fodor, G. and LePore, E., 1992. *Holism: A Shoppers Guide*. Cambridge: Blackwell.
- Fodor, G. and Lepore, E. (eds.), 1993. *Holism: A Consumers Update*. Amsterdam: Rodopi.
- Guttenplan, S. 1994, "Holism," in Guttenplan, (ed). *A Companion to the Philosophy of Mind*. Cambridge: Blackwell, 1994.
- Harman, G. 1973. *Thought*. Princeton: Princeton UP.
- Harman, G. 1993. "Meaning Holism Defended," in Fodor and Lepore 1993.
- Jackman, H. 1996. *Semantic Norms and Temporal Externalism*. Ph.D. Thesis, University of Pittsburgh.
- Jackman, H. 1999. "Charity, Self-Interpretation, and Belief," MS University of Toledo.
- Lormand, E. 1996, "How to be a Meaning Holist" *The Journal of Philosophy* Vol. XCIII, n.2. pp. 51-73.
- Margolis, E. and Laurence, S. 1998, "Multiple Meanings and the Stability of Content", *The Journal of Philosophy* Vol. XCV, n.5. pp. 255-63.
- McDowell, J. 1986. "Singular Thought and the Extent of Inner Space," in Pettit & McDowell (eds). *Subject, Thought, and Context*. New York: OUP, 1986.
- Sellars, W. 1974. "Meaning as Functional Classification," *Synthese* 27, pp. 417-37.

Quine, W.V.O. 1951. "Two Dogmas of Empiricism," in Quine, W.V.O. *From a Logical Point of View*. Cambridge: Harvard University Press, 1953.

Wilson, N. L. 1959. "Substance Without Substrata," *Review of Metaphysics* 12, pp. 521-39.