How to Argue Against (Some) Theories Of Content[*]

Michael V. Antony
Department of Philosophy
University of Haifa
Haifa 31905, Israel
Email: antony@research.haifa.ac.il

Abstract. An argument is offered against three naturalistic theories of intentional content: causal-covariation theories, teleological theories, and certain versions of conceptual role semantics. The strategy involves focusing on a normative problem regarding the practice of associating content expressions (e.g., that-clauses) with internal entities (states, symbol structures, etc.). The problem can be expressed thus: Which content expressions are the *right* ones to associate with internal entities? I argue, first, that an empirical solution to this problem—what I call the *Normative Problem*—will follow naturally from a descriptive-explanatory account of the practice of associating content expressions with internal entities; and second, that the empirical solution will be accepted and adopted within cognitive science. Naturalistic theories of content also entail solutions to the Normative Problem, and such theories are shown to be false by showing that their solutions to the Normative Problem are inconsistent with the empirical solution coming out of cognitive science.

In this paper I present an argument against three broad classes of naturalistic theories of intentional content for mental representations: causal-covariation theories, teleological theories, and certain versions of conceptual role semantics.[1] It would perhaps be more accurate to call what I present a *sketch* of an argument, since there are a number of points that demand more elaboration and defense than I can provide here. My main aim, however, is just to make plausible that there is a convincing argument against the three theories of content along the lines I suggest. In addition, although I confine myself to the three theories, I believe the basic strategy of my argument applies to any naturalistic theory of intentional content (hence the parenthetical 'some' in the title), and possibly to philosophical theories of meaning quite generally. I hope to provide the fuller story in a longer work.

1. Terminology

I begin with a few words about mental representations and naturalistic theories of content. I construe the concept *mental representation* broadly to include any entity that is internal[2] to a cognitive system (e.g., a state, structure, property, process, etc.), and satisfies the following two conditions:

---

[1] See, e.g., Stampe 1977, Dretske 1981, 1988, and Fodor 1987, 1990 on covariation theories; Milklikan 1984 and Papineau 1986 on teleological theories; and Block 1986 and Harman 1987 on conceptual role theories. Useful introductory discussions of the three theories are (respectively) McGlaughlin and Rey 1998, Godfrey-Smith 1998, and Block 1998.

[2] Externalism is not here being challenged. If externalism is true, representations can still be internal to cognitive systems in the sense in which mosquito bites or sunburns can be on arms.

(1) The entity is posited, or presupposed, in cognitive scientific explanatory practices.

(2) The entity is characterizable by means of the intentional idiom.

By 'the intentional idiom' I have in mind language about such internal entities that ostensibly makes reference to their meanings or intentional contents. Here are some examples of mental representations characterized in the intentional idiom:

- A belief that the shape is square
- A restaurant script
- A visual image of a stormy lake
- A bird concept
- A distributed representation of coffee

Each instance contains an expression for an internal entity ('belief', 'script', 'visual image', 'concept', 'distributed representation') and an expression meant to provide the entity's meaning or content ('that the shape is square', 'restaurant', 'a stormy lake', 'bird', 'coffee'). I call the latter expressions *content expressions*. Mental representations, then, are internal entities posited or presupposed by cognitive scientists that are (often) referred to, in part, by way of content expressions.

By a *theory of content* for mental representations I have in mind a theory for which the following two points hold:

(1) The theory presupposes that content expressions refer to (or contribute toward referring to[3]) properties of the internal entities mentioned above. For example, 'of a stormy lake' in the above example refers to a property of the internal entity referred to by 'visual image'.

(2) The theory attempts to say, in a systematic way, something substantive about the nature of those properties of the internal entities.

A constraint often placed upon any theory of content is that it be *naturalistic*, which is tantamount to the requirement that any account of content-properties fit within a broadly physicalist world view. A theory of content, however, must also specify how distinct content-properties relate to distinct features of the physical world. To that end, attempts are made to reduce content-properties to properties that are unproblematically physical.[4] According to causal covariation theories, for example, a state is properly characterizable with the content expression 'stormy lake' if that state causally covaries in the appropriate manner with the property *stormy lakehood*. That satisfies the naturalism constraint since both causal covariation and

---

[3] The reason for the qualification is that if an internal entity's having a certain content is a relational property (e.g., covariation with property *P*), then it might be claimed that the content expression refers to the relatum to which the internal entity is related (e.g., property *P*) rather than the property of having such-and-such a content. In that case the content expression would merely *contribute* toward referring to the property of having such-and-such a content. Such niceties will be ignored in what follows.
[4] The argument of this paper also applies to Fodor's (1987,1990) theory which merely aims to provide sufficient conditions for intentionality.

stormy lakehood are unproblematically physical; and different internal entities are assigned different contents in virtue of covarying with different properties.

2. The Normative Problem and the Descriptive Problem

In arguing against the three naturalistic theories of content (hereafter I drop 'naturalistic'), I first focus on a normative problem regarding the cognitive-scientific practice of ascribing content expressions to, or associating them with, internal entities.[5] The problem can be expressed thus: Which ascriptions of content expressions to which internal entities should be sanctioned within cognitive scientific practice? Or more simply: Which content expressions are the *right* ones to ascribe to internal entities? This problem makes itself felt most strongly where systems under study deviate to a significant degree from normal adult humans (for example, brain-damaged adults, pre-linguistic children, non-human animals, robots). In studying such systems, practitioners often find themselves unsure about which content expression to associate with a given internal state. Let us call this problem about the correct ascription of content expressions to internal entities *the Normative Problem*.

At the most general level, my argument against the three theories of content is this:

(1) As a matter of empirical fact, the Normative Problem will be solved from within cognitive science, and that solution will not appeal to a philosophical theory of content.

(2) This "empirical solution" to the Normative Problem, as I shall call it, will be accepted and adopted within cognitive science, since it will be empirically and theoretically grounded.

(3) Each of the three theories of content entails a solution to the Normative Problem. Consequently each theory's only hope is for its solution to be consistent with the empirical solution to the Normative Problem.

(4) The solution to the Normative Problem entailed by each of the theories of content is inconsistent with the empirical solution to the Normative Problem.

(C) Therefore, each of the three theories of content is false.

In what follows, I focus mainly on premise (1), the claim that there will be an empirical solution to the Normative Problem. Concerning (2), the claim that the empirical solution to the Normative Problem will be adopted, I shall say a bit below. Regarding (3), the premise that each of the theories of content entails a solution to the Normative Problem, I shall say only this: covariation, teleological, and conceptual-role theories all aim to provide truth conditions for content ascriptions. Sanctioned or correct ascriptions, therefore, are by and large those that are *true;*[6] and which are true

---

[5] My use of 'ascribe' ('ascription', etc.) is somewhat idiosyncratic. In addition to speaking of ascriptions of mental states (e.g., beliefs) to cognitive systems, I speak of ascriptions of content expressions to internal entities (e.g., symbols), and also to cognitive systems. 'To ascribe', in this sense, is to *associate with* or *assign to*.

[6] The point is discussed further in section 9.

is determined by the details of the theory. On a covariation account, for example, an ascription of the content expression 'dog' to an internal entity is true (and hence sanctioned) if that entity covaries with the property *doghood*. Premise (4), finally, will be defended toward the end of the paper. I turn now to premise (1).

How will a solution to the Normative Problem arise within cognitive science? I believe a solution will be a natural byproduct of a solution to another problem in cognitive science—that of providing a descriptive-explanatory account of the practice of ascribing content expressions to cognitive systems. To bring this out more clearly, consider a typical situation in which content expressions are ascribed. There will be a *speaker* who produces an ascription, a *hearer* for whom the production is intended, and a *cognitive system* (S) to which the content expression is ascribed. For example, a speaker might say to a hearer concerning some system S: 'S thinks that the squirrel is behind the tree'. The hearer will then process the utterance and come to think of S in a new way. Now clearly there are many aspects of such interactions among speakers and hearers that cognitive science will want to describe and explain. Regarding speakers, for example, an account will seek to answer questions such as these:

(1) How does the speaker represent the system before producing the ascription?

(2) How does the speaker represent the hearer's representation of the system before producing the ascription?

(3) What changes in the hearer's representation of the system does the speaker wish to bring about by his ascription?

(4) Why does the speaker wish to bring about those changes in the hearer?

(5) How does the speaker choose a content expression to effect those changes?

Concerning the hearer there are questions such as these:

(6) How does the hearer represent the system before processing the speaker's ascription?

(7) How does the hearer represent the system after having processed the ascription?

(8) How does the processing of the content expression cause the representational change?

An account that aims to answer such questions is what I have in mind by a descriptive-explanatory account of the practice of appealing to content expressions.[7] Let us call the problem of providing such an account within cognitive science *the Descriptive Problem*, to contrast it with the Normative Problem. Now my claim is that an empirical solution to the Normative Problem will follow naturally from a cognitive-scientific solution to the Descriptive Problem. That, I submit, is how the Normative Problem will be solved from within cognitive science.

---

[7] For one approach see Sperber and Wilson 1986.

3. How *Could* a Solution to the Descriptive Problem lead to a Solution to the Normative Problem?

I now want to provide a sense of how a solution to the Descriptive Problem *could* give rise to a solution to the Normative Problem. In the next section I shall sketch the story in more detail.

Suppose that as a result of cognitive scientists describing the practice of ascribing content expressions to internal entities, they came to see that that practice is a *good* one, given the aims of the science. What must be imagined here is that a descriptive-explanatory account of what practitioners do would show why their practice is good for doing cognitive science. Such an account, for example, might show that ascribing content expressions is a highly flexible and efficient means of transmitting information about complex cognitive systems, it might show in what this flexibility and efficiency consists, and also that it is hard to imagine how else that might be accomplished. Under such circumstances, cognitive scientists might judge that their practice of ascribing content expressions is a good one, perhaps even indispensable. The overall practice would thus be justified, in a global way, from within cognitive science. Such justification would not be based on an epistemological theory of justification, but on practitioners' beliefs about the aims of their science, along with an empirical theory demonstrating how their practice furnishes them with a powerful, possibly indispensable tool for pursuing those aims.

Thus far we have imagined that from a solution to the Descriptive Problem, global justification of the practice of ascribing content expressions could emerge. However global justification would not in itself amount to a solution to the Normative Problem, for the Normative Problem concerns *particular* ascriptions of content expressions, not the practice as a whole. One way a solution to the Normative Problem could emerge is from a combination of the following three factors: (1) global justification of the practice, (2) the details of the solution to the Descriptive Problem, and (3) the principle that if the practice as a whole is deemed good, then individual ascriptions are good to the extent that they conform with the practice. What "conforming to the practice" amounts to would be given, in turn, by the details of the solution to the Descriptive Problem. That, essentially, is how I believe the Normative Problem could—and will—be solved within cognitive science. Notice that no naturalistic theory of content is required for the solution.

A few words can be offered at this point in defense of premise (2), the claim that the empirical solution to the Normative Problem will be adopted, since it will be empirically and theoretically grounded. The idea is that if an empirical solution to the Normative Problem were to arise in the manner just described (by way of an empirical story detailing the strengths of the practice, and hence the strengths of producing individual ascriptions in accordance with the practice) then, first, the solution would be adopted within cognitive science, and second, no conflicting non-empirically-grounded philosophical theory would displace it. I assume that the first point—that the solution would be adopted among cognitive scientists—is obvious. And concerning the second point, the solution would not be displaced because the ascriptions sanctioned by the empirical solution, by assumption, would have been

shown to advance the aims of the science better than any alternative, including any coming from philosophy. This is all I shall say here in support of premise (2).[8]

By now an objection will have occurred to some readers. It may seem that my strategy for arguing against theories of content is hopeless, since it depends on some as yet non-existing empirical theory turning out a particular way. But surely we cannot know how future scientific theory will turn out? While I agree we can know little about how psychology theory will develop, I believe we can make some reasonable projections about the general features that a descriptive-explanatory account of content-expression ascription is likely to have. Such features, moreover, are sufficient to see that the Normative Problem *will* be solved, and also that the solution will be inconsistent with the solutions entailed by the three theories of content. Now it may turn out that these speculative claims about future theory will be shown to be deeply mistaken. In that case my argument will collapse. The fact that my argument has empirical implications, however, I count as a point in its favor.

In the remainder of the paper I sketch the general features of the descriptive story, the solution to the Normative Problem that arises from it, and the argument against the three theories of content.

4. A Sketch of the Descriptive Account

I center my discussion of the Descriptive Problem on the typical situation mentioned above involving a speaker, hearer, and system S, in which content expressions are ascribed. I make three main points: the first about the speaker, the second about the hearer, and the third about the speaker again. From these three points a general picture of the practice emerges, from which the idea of global justification is developed.

The first point regarding the speaker is as follows. Typically, in ascribing content expressions to cognitive systems, speakers wish to get across to hearers *fine-grained psychological details* of cognitive systems.[9] Somewhat more accurately, speakers have in mind fine-grained psychological details of systems, and they wish to bring it about that their hearers, after processing their utterance, have the same (or sufficiently similar) fine-grained details in mind concerning the system. This requires some clarification.

By 'fineness of grain' I mean *amount of psychological detail:* the more detail in a psychological description, the finer the grain. I assume that with respect to virtually any mental representation, there is an indefinite amount of psychological detail that can be expressed regarding it, detail which could be relevant in some or other explanatory context. Such detail might concern the internal structure of the representation (assuming it has structure), or its external relations (e.g., as given by its conceptual role, causal relations to the environment, etc.).[10] To illustrate, consider

---

[8] Additional support comes from the details of the empirical solution to the Normative Problem, some of which I shall provide below, and the fact that philosophical theorizing rarely dislodges empirically grounded scientific results.

[9] For related claims, see e.g. Sperber and Wilson 1986, Loar 1988, and Travis 1997.

[10] Notice that nothing is being assumed here about representations' essential properties, identity conditions, etc.

some expressions involving increasing amounts of psychological detail regarding some token visual image—that is, increasingly finer-grained descriptions of it:

- image of a stormy lake
- image of a stormy, dark green lake
- image of a stormy, dark green lake, possessing such and such structural features
- image of a stormy, dark green lake, possessing such and such structural features, which typically causes psychological process X.

In principle, one could go on indefinitely about such an image, in ever increasing amounts of psychological detail. And the same holds for any other mental representation.

My claim about the speaker was that speakers typically wish to convey to hearers fine-grained psychological details of systems. Since fineness of grain is a relative matter, the question arises: How fine-grained? My answer is: Fine-grained relative to the amount of detail explicitly mentioned in the content expression. My point about speakers in a more unpacked form, then, is this:

> When a speaker has in mind psychological details about a system, a representation of which he wishes to bring about in a hearer, those details typically (or at least often) are finer-grained than the details explicitly mentioned in the content expression the speaker chooses for the task.

Take the ascription 'S thinks that the squirrel is behind the tree'. The claim is that a speaker typically will have in mind certain details regarding S's *particular way* of representing squirrels, trees, what a squirrel's being behind a tree is likely to involve, etc.; and that the speaker will want to convey some of those details to the hearer by means of the ascription.

A way to begin to see this is to imagine yourself ascribing the same content expression to different types of systems, and then reflect on what, intuitively, the ascription would be saying about each system—what the psychological "cash value," as it were, would be of holding the ascription true of each system. Imagine, for example, ascribing 'S thinks that the squirrel is behind the tree' to an adult human, then to a three-year old toddler, then to a dog, then to an AI program running in a computer. Consider in each case the range and nature of the psychological details (types of perceptual or cognitive states, dispositions, etc.) you would be inclined to project onto the system, perhaps with varying degrees of confidence. In performing this thought experiment myself, I find that the psychological details I project differ substantially from system to system. Another way to see the point is to imagine ascribing the same content expression—say, 'S fears the war will be disastrous'—to systems about which you have varying degrees of knowledge—for example, to a close friend whose political views you know well, and to someone about whom you know little. Here there seems to be a much richer projection of detail (e.g., regarding the specific sort of disaster feared) onto the close friend than the stranger.

If one gets and accepts the intuition I have been trying to generate, it follows that in processing the same content expression with respect to different systems, at

least sometimes different fine-grained details are projected. If that is true, however, we should expect it to be equally true of both speakers who produce ascriptions and hearers who comprehend them. There is no reason it should hold for only one of them. But from there it is a short step to the idea that fine-grained details about systems are what speakers *want* to get across to hearers (unconsciously, no doubt). That, in any event, is my first point about speakers.[11]

> The second point concerns the hearer, and is this:

> With respect to the wide range of possible hearers who are competent in the language, and with respect to some content-ascription a speaker produces regarding a system S, if each hearer were to process that content-ascription, there would be significant differences among them in the fine-grained details they project onto S.

If we accept the first point discussed above—that speakers typically want to convey fine-grained details to hearers—this second point generates a puzzle. The puzzle is that the practice seems to work far better than that would suggest, since speakers typically seem to succeed in conveying what they want. This puzzle will be resolved in the discussion of the third point below.

For now I want to show why different hearers would end up with widely differing fine-grained representations of a system S, if they were to process the same content expression. This can be seen by attending to two respects in which individuals differ from each other psychologically, and then noting how differences among individuals in those respects would determine fine-grained differences in such individuals' representations of S. The two respects are: (1) individuals' knowledge, beliefs, experience, etc. concerning S—what I shall call 'system knowledge'; and (2) individuals' conceptual roles for the mental representations corresponding to the words in the content expression. I discuss these points in turn.

In arguing that speakers wish to convey fine-grained psychological details to hearers, I tried to evoke the intuition that, in ascribing a content expression to different systems (adults, babies, dogs, computers), different fine-grained representations are projected onto those systems. This same intuition shows that distinct individuals who differ in their knowledge or beliefs regarding a system will generate different fine-grained representations when processing the same content expression. For when we considered different systems, our different projections of fine-grained details resulted from our having in mind different "system-representations" when applying the content expression. But individuals who differ in their knowledge or beliefs concerning a system also differ in the system-representations they bring to the task of processing ascriptions to the system. Indeed, they do so trivially. Consequently, they too will typically differ in the fine-grained representations they project onto systems. Now since individuals differ enormously in

---

[11] My strategy for showing that speakers wish to convey fine-grained details regarding systems has been to show that, across *different* systems, speakers may use the same content expression to convey different fine-grained details. The same point can be made with a single system. Though I cannot develop the point here, the idea is that depending on a speaker's *purposes* for making an ascription, a speaker may wish on different occasions to convey different details about S using the same content expression.

their knowledge and beliefs concerning systems, it follows that there would be considerable differences in the fine-grained representations they would project onto any given system.

I shall say little about how different conceptual roles for the mental representations corresponding to the words in a content expression contribute to different fine-grained representations being projected onto systems. The point is fairly obvious. As people's conceptual roles for mental representations corresponding to 'squirrel' and 'tree', for example, differ, so will their semantic processing of the ascription 'S thinks that the squirrel is behind the tree', and thus, in many cases at least, so will the fine-grained representations they project onto S.

We see, then, that different hearers, all competent in the language, would end up with widely differing fine-grained representations of a system from processing the same content expression. Differences in system knowledge and conceptual role will have been largely responsible. But now what about the puzzle mentioned earlier—that speakers typically seem to succeed in getting across what they want? This puzzle suggests that there is something missing from our story. To locate the missing element we need to turn again to the speaker.

Our discussion of the features that play a role in determining the fine-grained representations one projects onto systems (system knowledge and conceptual role) suggests a solution to the puzzle. Suppose it were built into the practice of using the intentional idiom for speakers to know (tacitly) that hearers' system knowledge and conceptual roles affect the fine-grained representations of systems they end up with when processing content expressions. If that were so, then assuming our first point— that speakers want to get across fine-grained details about systems—we might expect speakers to attempt to assess hearers' system knowledge and conceptual roles in order to determine whether they, as they are, would generate the desired fine-grained details the speaker wishes to convey. If it appears that they would, then the speaker need only make the ascription. However, if the assessment is that they would not, we might expect the speaker to try to provide the needed information concerning the system and/or conceptual roles, along with the ascription of the content expression. Doing so would increase the likelihood of the hearer representing the desired fine-grained details.

Now I claim—and this is my third point—that something very much like this is what speakers do. And that is the solution to the puzzle. I also believe that this, or something very much like it, will be confirmed by future cognitive scientific theory. Here I can offer only a few examples in its defense.

Consider an individual of the sort described by Tyler Burge (1979) in his famous anti-individualism thought experiment. Burge describes someone who misunderstands the concept *arthritis*. The person believes falsely that one can have arthritis in the thigh—falsely, since arthritis is a disease of the joints. Burge argues that in spite of this confusion the person can nevertheless be correctly ascribed many propositional attitudes (de dicto) using 'arthritis'—for example, the belief that arthritis is painful. Now suppose we ascribe to such a person the belief that arthritis is painful. It is easy to imagine cases in which we would ascribe that belief without feeling any need to inform our audience about the person's conceptual confusion. It

9

may simply be irrelevant to our purposes on that occasion. However, there clearly also are contexts in which we *would* want to add this fact about the individual, for example, whenever it would be important that our audience be able to explain or predict the individual's thought or behavior related to the confusion.

Cases in which we provide additional information about systems, along with our intentional ascriptions to them, are legion. Another example would be an AI researcher intending to discuss certain representational structures and processes in some system, but before doing so sketches a number of general (or specific) features concerning the system. Or suppose I want to speak of a person whom I know well but you do not, and I want you to know various subtle details about his view of some particular war. I might preface my ascription 'S is afraid that the war will be disastrous' with a story about aspects of S's psychology, S's political outlook, etc. Of course, if I believe that you know S as well as I, I will offer the ascription on its own, trusting that you will represent the fine-grained features I want you to.

I believe that filling in extra information about systems to which content expressions are ascribed (when deemed necessary) is a central feature of the practice of using the intentional idiom. When filling in does not occur, I would claim, it is because it has been judged that the hearer already possesses the relevant information. Similar points hold concerning adjusting for hearers' conceptual roles, though I shall not discuss them here.

The above three points, then, comprise the basic facts about the practice (assuming they are facts) that I need for my argument against the three theories of content. From those facts, a general picture of the practice emerges. The details of that picture lead to global justification, to the solution of the Normative Problem, and to an argument against the three theories of content.

5. The Emerging Picture

What can we say about the practice on the basis of the three points? We have a picture according to which, for each speaker-hearer interaction concerning a system, different fine-grained details may be needed to account for what occurs psychologically, even when the same content expressions are ascribed to the same systems. For example, there will typically be idiosyncratic, fine-grained details regarding what the speaker wants to convey to the hearer, what the speaker believes regarding the hearer's system knowledge and conceptual roles, what the hearer's system knowledge and conceptual roles are in fact like, and so on. The practice is thus seen to work largely at a level of fine-grained, idiosyncratic psychological detail.

In contrast with this level at which fine-grained psychological differences are emphasized, common patterns are also apparent in the workings of the practice. For example, all speakers are seen as wanting to get across fine-grained details of systems to hearers, and, to that end, as attempting to "manipulate" hearers' representations to orient them to those fine-grained features. Speakers do this by attempting to choose appropriate content expressions, take into account hearers' system knowledge and conceptual roles, and fill in where deemed necessary to fine-tune the result as best as they can. A further feature of the practice that emerges is that there is considerable "messiness." Things do not work perfectly. Speakers succeed to various degrees in

conveying what they want. Error enters in at many points (for example, in speakers' beliefs about hearers' system knowledge and conceptual roles). But errors can also often be detected, and compensated for. Moreover, all of this in principle can be observed and described.

Notice that if I am right about future theory of the practice, the features I have described will also apply to the theorists providing the theory. That, of course, is just what we should expect of a cognitive scientific account of (aspects of) cognitive scientific practice. In this paper we have been playing the role of theorists theorizing about speakers, hearers, and systems. If what has been said about speakers' and hearers' processing of content-expressions is correct, then it also applies to us in our own thinking about the speakers and hearers in the above examples.

6. Global Justification

We can now say how global justification arises. One point that emerges from the descriptive account of the practice is that reference to fine-grained psychological details is unavoidable, since such details play a role in all stages of processing. These fine-grained details, varying across individuals and conversational contexts, provide a "background" against which the overall practice must be understood. Seen against that background, the practice seems remarkably well suited to it. As we have seen, built into the practice are ways of compensating for fine-grained differences among individuals. However, the practice also *exploits the numerous fine-grained similarities that exist among individuals*. So far we have focused on psychological differences, but there also are countless fine-grained psychological similarities among individuals, based on common genetic factors, learning histories, etc. The number and kinds of psychological similarities will differ for each pair of individuals, but in general there will be numerous similarities for any pair. Now in compensating for differences, speakers "fine-tune" what they think will arise in hearers' heads. What does *not* get fine-tuned, is what the speaker assumes he or she has in common with the hearer (or assumes is irrelevant). It is this rich, common, mental structure that is exploited in the practice.

A nice way of seeing what this exploitation comes to is to imagine yourself talking to a near physical duplicate of yourself, say a creature which five minutes ago was your molecule-for-molecule twin. Imagine saying to your twin that a friend of yours is afraid the war will be disastrous. Using the imagery of semantic networks and "spreading activation,"[12] we could describe what happens thus: The words you utter activate (unconsciously, of course) conceptual nodes in your twin's "psychological space," and this activation spreads out and activates countless psychologically associated nodes. These correspond to the idiosyncratic ways in which fear and war and disasters are conceptualized, to beliefs and experiences regarding your friend, etc. All of this is activated in your twin's brain as a result of having processed your ascription. Now since your hearer is your twin, the very same details that are activated in your brain when you process the content expression are activated in your twin's, and at the finest level of grain. You thus cannot fail to get across to your twin exactly what you have in mind about your friend's fear of the war, in all its richness of detail—and all by emitting a few simple sounds. In accomplishing this feat you will

---

[12] See, e.g., Collins and Loftus 1975, Anderson 1983.

have exploited the fact that the words in the content expression have the same psychological effects in your twin's brain as in yours; you will have exploited the perfect psychological similarity between you. Now, of course, with any actual non-twin there will be many differences, but your words will still activate psychological structures very similar to your own, and it will be possible to exploit those similarities.

Exploiting psychological similarities means that a lot can get done with a few sounds or marks. The practice is thus extremely efficient. Because of the (tacit) awareness speakers have of psychological differences among hearers, and the nature of the fine-tuning that aims to compensate for such differences, the practice also displays great flexibility. For the same content expression can do different jobs for different hearers and systems, in the sense of orienting different hearers to different fine-grained details of systems.

I hope this is enough to see how, if there were a confirmed descriptive account along the lines I have been describing, the practice as a whole could be deemed "good" from within cognitive science. In this sense the practice would be globally justified.[13] It would be seen as a powerful tool—in the sense of being highly efficient and flexible—for communicating subtle, fine-grained psychological features of complex cognitive systems. The cost for this power would be a degree of messiness and error, but it would be of the sort that frequently could be detected and remedied. Now I claim there *will* be a descriptive story containing these features, and that the practice will be globally justified in much this way.

7. The Empirical Solution of the Normative Problem

We said earlier that, given the details of the solution to the Descriptive Problem and global justification, the Normative Problem gets solved according to the following principle: If the practice as a whole is deemed good, then individual ascriptions are good to the extent that they conform with the practice. Sanctioned individual ascriptions, then, are those that accord with the practice; and it is from the details of the descriptive story that we are to understand what "according with the practice" amounts to.

To argue against the three theories of content, I must show that the empirical solution to the Normative Problem is inconsistent with the solutions to the Normative Problem entailed by those theories of content. I need not produce a "complete" solution to the Normative Problem, nor could I. I need only say enough about sanctioned content ascriptions according to the empirical solution to generate inconsistencies with the three theories of content. Saying enough to generate such inconsistencies is fairly straightforward. Through one central insight into the practice, arrived at by way of the descriptive account, it can be seen that a wide range of ascriptions are sanctioned that are inconsistent with those sanctioned by the three theories of content. The insight concerns the flexibility mentioned above that arises from the possibility of manipulating hearers' system knowledge and conceptual roles to fine-tune hearers' representations of systems. Let me explain.

---

[13] Though I shall not argue the point here, I believe the practice would also be judged indispensable, thus ruling out eliminativism regarding the intentional idiom.

I appeal again to the imagery of semantic networks and spreading activation. According to the descriptive story, the content-expression which a speaker produces activates related nodes in the hearer's semantic network, and that activation spreads to countless fine-grained details in the hearer's psychological space. We have discussed how the hearer's system knowledge effects the representation of the system the hearer ends up with. Here is how that might work. The content expression activates elements in the hearer's psychological space, and then the hearer's system knowledge is applied to "prune down" this overall set of semantic elements to just those that are applicable to the system, given the type of system it is. Think, for example, of some of the associations the word 'squirrel' generates in your own mind. They might include, being an animal, being a rodent, having a bushy tail, having such-and-such prototypical visual appearance, burying acorns before winter, scurrying up trees, etc. Now consider the ascription 'S thinks that the squirrel is behind the tree'. It should be clear how the overall stock of associations generated in your mind by that content expression might be pruned down based on your beliefs about S. If S is a five-year old child, the fact that squirrels are rodents will likely be ruled out as inapplicable. If S has no visual system (say, because S is a computer), or is a human blind from birth, a squirrel's visual appearance (or at least aspects of it) may be ruled out as inapplicable. And so on. In this way your system knowledge of S is applied to eliminate elements in your psychological space that have been activated by 'squirrel' and 'tree' in order to generate an interpretation of the content expression that is appropriate to S.

Now the crucial point is this. The speaker wants to ensure that the hearer has the required system knowledge for pruning down the hearer's representation to the fine-grained details the speaker wishes to convey. If the speaker thinks the hearer lacks the required system knowledge, the speaker will supply it. But notice: *there are no significant restrictions on what or how much pruning down can legitimately be done*. Indeed, the descriptive story shows that what in fact gets pruned down varies in an immense number of ways. Not only is "pruning down" used to enable the same content expressions to be applicable to adult humans, children, animals, and machines, but also to an indefinite number of variations within each of those categories. In humans there can be countless levels of competence or skill, differences in knowledge, belief, experience, perceptual or motor abilities, and so on. And similarly with children, animals, and machines.

We thus have the outlines of a single, unified story that accounts for what goes on in all such cases of content ascription. No cases are privileged. It follows that they all have the same status according to the descriptive story—namely, *they are all equally sanctioned*. We thus have a significant part of the solution to the Normative Problem.

8. The Argument Against Theories of Content

We can now, finally, argue against the three theories of content, and rather quickly. Consider first covariation and teleological theories. The main problem I wish to bring out concerning those theories is that they fail to sanction a great many ascriptions that are clearly sanctioned by the empirical solution to the Normative Problem. Their solutions to the Normative Problem are thus inconsistent with the empirical solution. For instance, both theories require that for a content ascription to be true of a system

(and hence sanctioned), the system must bear some naturalistic relation to the properties expressed in the content expression. With covariation accounts, as was mentioned in section 1, internal symbols must causally covary with such properties. With teleological accounts, the representation (or system) must have the *purpose* or *function* (determined by evolution, or by human design in the case of artificial systems) of bearing some relation to the properties.[14] However, according to the nature of the practice, as it has been described, it is perfectly legitimate to ascribe content expressions to systems that have no sensory-motor systems, or extremely limited ones (and so involve little or no covariation), and were not designed to have any. A computer running a sophisticated AI program with no programming for sensory-motor systems is a case in point. And regarding systems that possess sensory-motor systems, it is in keeping with the empirically-determined nature of the practice to ascribe content expressions to such systems even when the internal symbols *systematically* fail to covary with the properties mentioned in the content expression. (Obvious cases are many common ascriptions to babies, animals, cognitively-impaired adult humans, etc.) Moreover, systems need not have been designed at all: it is perfectly in keeping with the practice to ascribe content expressions to bizarre quantum accidents or Davidsonian "Swampman" examples.[15] Ascriptions to all of these systems simply involve—as always—the pruning away of different features to generate an appropriate representation regarding the system. It just happens that what get pruned away in these examples are the covarying and teleological properties that the theories of content insist are essential.

Turning now to conceptual role accounts, consider first those requiring causal relations to properties outside of the head. Such accounts treat meaning either as having a single factor or two factors. On the single factor view, conceptual roles ultimately are external—they extend out to properties in the system's environment. On two-factor accounts, the role of a representation *within* a system (in a system's brain, say) is distinguished from the representation's *external* relations to the environment.[16] Both one- and two-factor theories can immediately be rejected for reasons already rehearsed: content expressions properly ascribable to adult humans are equally sanctioned for computers running AI programs with no programming for sensory-motor systems. But with no sensory-motor system there can be no (relevant) relations to external features of the envinronment. Regarding conceptual role theories involving only an internal factor, we can distinguish two kinds. First, there are those on which the meaning-constitutive roles are few in number, comprising "a select set," such as those that determine analyticities, or Peacocke's (1992) "primitively compelling inferences," etc. I shall not argue against such theories in this paper, aside from pointing out that no good reasons have been offered for thinking such roles can be provided for *any* intentional contents (aside, perhaps, from certain logical concepts like *and*, *negation*, etc.). Second, there are internal conceptual role theories according

---

[14] The precise details of the relation vary across different versions of the teleological theory, but are unimportant for our pourposes. See Godfrey-Smith 1998 and Neander 2004 for discussion.

[15] Swampman examples involve a physical duplicate of a human being that comes into existence instantaneously and coincidentally (emerging from a swamp by some quantum accident, for example). Although physically identical to a human, Swampman lacks any personal or evolutionary history. Swampman examples thus create trouble for teleological theories which most naturally treat "swamp-creatures" as lacking intentional states entirely. See Davidson 1987, and Neander 2004 for a general discussion.

[16] See e.g. Block 1986 and Harman 1987 on two-factor and single-factor theories, respectively; and Block 1998 for an overview.

to which the number of meaning-constitutive roles is far greater, extending even to *total* (internal) conceptual role.[17] But as should be clear by now, theories of this sort are ruled out by the wide range of sanctioned uses that differ from any *specific*, extensive conceptual role that might be specified.

9. An Objection

I close by considering an objection that will have occurred to many readers. In arguing that many ascriptions sanctioned by our practice are not sanctioned by the three theories of content, I have assumed that the theories of content sanction only ascriptions they count as *true* (in virtue of representations suitably covarying, having the right purpose or function, or conceptual role). However that is open to dispute. Why, for example, could there not be sanctioned ascriptions, in the sense of pragmatically appropriate ascriptions, that are strictly speaking false? After all, everyday language is replete with cases in which literally false sentences are legitimately employed to achieve sundry effects (irony, metaphor, Gricean implicatures,[18] etc.). Similarly, the objection goes, from the fact that it may be *useful* to ascribe the same content expression in different sorts of circumstances, or different content expressions in the same circumstances, it does not follow that all or most such ascriptions are *true*. Thus it can be acknowledged that content expressions are ascribed much as I have claimed in this paper, while at the same time maintaining that many or most of those ascriptions are false—from which it follows that inconsistencies between ascriptions sanctioned by our practice and those sanctioned by the theories of content are insufficient for inferring the falsity of the theories.

My reply is as follows. Although it is true that the falsity of the theories of content is not entailed by the inconsistency between the sanctioned ascriptions of our practice and those of the theories, fortunately nothing as strong as entailment is required. That is because to make the objection work, a *theory* is needed that explains, or at least illuminates, both the existence and role of the countless purportedly false ascriptions within our practice. In the absence of such a theory the objection is blatantly ad hoc. For comparison, notice that there is a theory of conversational implicature which explains how false utterances can play a role in a hearer's computing what a speaker intends to communicate.[19] But there is no comparable theory to support the objection under consideration, not even a hint a what such a theory might look like. So the response is ad hoc: there is no reason to suppose that numerous ascriptions sanctioned by our practice are false, aside from the fact that their falsity is required by the theories of content. Indeed, the ascriptions are not even *intuitively* false (interpreted literally), in contrast with irony, metaphor, many Gricean implicatures, etc. I conclude that the objection fails. That does not mean that there could not be a theory of the sort required, just that there is currently no non-ad-hoc reason to believe there could be such a theory. It is thus most rational, at least for the time being, to conclude that the three naturalistic theories of content discussed in this paper are false.

Bibliography

---

[17] See e.g. Block 1986, p. 628
[18] See Grice 1989.
[19] See, e.g., Grice 1989.

Anderson, J. (1983). *The Architecture of Cognition*. Cambridge, MA: Harvard University Press.

Block, N. (1986). "Advertisement for a Semantics for Psychology." In P. French, T. Uehling, & H. Wettstein (Eds.), *Midwest Studies in Philosophy, Vol. 10* . Minneapolis, MN: University of Minnesota Press.

Block, N. (1998). "Semantics, Conceptual Role," in E. Craig (Ed.), *Routledge Encyclopedia of Philosophy*. London: Routledge.

Burge, T. (1979). "Individualism and the Mental." In P. French, T. Euhling, & H. Wettstein (Eds.), *Midwest Studies in Philosophy* (Vol. 4,). Minneapolis: University of Minnesota Press.

Collins, A. M., & Loftus, E. F. (1975). "A Spreading Activation Theory of Semantic Processing." *Psychological Review* **82**, 407-428.

Davidson, D. (1987). "Knowing One's Own Mind," in *Proceedings and Addresses of the American Philosophical Association* **60**, 441-58.

Dretske, F. (1981). *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.

Dretske, F. (1988). *Explaining Behavior*. Cambridge, MA: MIT Press.

Fodor, J. (1987). *Psychosemantics*. Cambridge, MA: MIT Press.

Fodor, J. (1990). *A Theory of Content and Other Essays*. Cambridge, MA: MIT Press.

Godfrey-Smith, P. (1998). "Semantics, Teleological," in E. Craig (Ed.), *Routledge Encyclopedia of Philosophy*. London: Routledge.

Grice, P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Harman, G. (1987). "(Non-Solipsistic) Conceptual Role Semantics." In E. Lepore (Ed.), *New Directions in Semantics* . London: Academic Press.

Loar, B. (1988). "Social Content and Psychological Content." In R. Grimm & D. Merrill (Eds.), *Contents of Thought* (pp. 99-110). Tucson: University of Arizona Press.

McGlaughlin, B. and Rey, G. (1998). "Semantics, Informational," in E. Craig (Ed.), *Routledge Encyclopedia of Philosophy*. London: Routledge.

Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT Press.

Neander, K. (2004), "Teleological Theories of Mental Content," *The Stanford Encyclopedia of Philosophy (Summer 2004 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2004/entries/content-teleological/>.

Peacocke, C. (1992). *A Study of Concepts*. Cambridge, MA: MIT Press.

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

Sperber, D., & Wilson, D. (1986). *Relevance*. Cambridge, MA: Harvard University Press.

Stampe, D. (1977). "Toward a Causal Theory of Linguistic Representation." In P. French, T. Uehling, & H. Wettstein (Eds.), *Midwest Studies in Philosophy, Vol. 2* . Minneapolis, MN: University of Minnesota Press.

Travis, C. (1997). "Pragmatics." In B. Hale & C. Wright (Eds.), *A Companion to the Philosophy of Language* (pp. 87-107). Oxford: Blackwell.