
Conceptual Structure and the Individuation of Content

Philosophical Perspectives, 1995, pp. 401-426.

Derk Pereboom, University of Vermont

Current attempts to understand psychological content divide into two families of views. According to externalist accounts such as those advanced by Tyler Burge and Ruth Millikan, psychological content does not supervene on the physical features of the individual subject, but is fixed partially by the nature of the world external to her.¹ In the rival functional role theories developed by Ned Block and Brian Loar, content does supervene on the physical features of the individual, and is, in addition, determined solely by the role it plays in the causal network of an individual's sensations, behavior, and mental states.² Over the past fifteen years, criticism of these two types of views has often focussed on their capacity to individuate content in an acceptable way, and both seem to be deficient in this respect.

A satisfactory theory must individuate content at a level that captures the generalizations in psychological explanation we want to make. Functional role theory allows for an extremely fine-grained specification of content, since any difference in associated belief can make for a difference in content. However, this type of theory seems to provide no principled criteria for individuating content in a way coarse-grained enough to explain how it might remain the same despite differences in associated beliefs, and thus content determined by functional role would seem to be incapable of sustaining the right psychological generalizations.³ Externalist theories, by contrast, furnish anchors for content by linking the cognitive value of concepts to entities in the external world, entities which retain their identity through variation in associated beliefs. But critics have argued that in some cases externalist theories nevertheless cut content too finely, and furthermore, that on other occasions externalist individuation is too coarse.⁴ In some cases, contents that externalist theories distinguish need to function in the same type of role for explaining inference and behavior, whereas in other situations contents that they render equivalent

must account for significantly different patterns of inference and behavior.

I believe that the force of these criticisms persists, despite attempts to mitigate their impact, but that there is a strategy for individuating content that might well not suffer from deficiencies in fineness or coarseness of grain. This approach takes its cue from the individuation of natural kinds in sciences such as chemistry and biology. In these sciences, structural features of entities play a prominent role in characterizations of the natural kinds of which they are members. Recent studies in cognitive psychology suggest that psychological content also has various structural features. I shall argue that these features can be appropriated for the individuation of content, and furthermore, that such individuation has a reasonable prospect of capturing fundamental intuitions about psychological explanation, and of providing principled criteria for identity of content and concept.

I

In psychological explanation of inference and behavior, propositional attitudes such as beliefs, desires, and wishes have a key role. Part of the explanatory power of a propositional attitude is attributable to its being the kind of attitude it is, for example, its being a belief rather than a desire or a wish. The remaining explanatory role of an attitude, by one taxonomy, is played by its content, for example, its being a belief that it is raining, rather than that it is not raining. But this taxonomy is not universal. In David Kaplan's semantic theory, for example, 'content' refers to the proposition expressed by the 'that' clause, where the proposition Socrates is wise is the same proposition as he is wise, given that 'Socrates' and 'he' are coreferential.⁵ In his taxonomy, besides content and type of attitude, a propositional attitude includes character, which includes features of how a subject represents a proposition. In Jerry Fodor's system, besides type of attitude and content in Kaplan's sense, propositional attitudes include vehicle, the symbol by means of which one's contents are represented, and functional role, the causal role a vehicle plays in a system of inference patterns.⁶ A discussion of content, therefore, requires that one set out the definition of

the notion one is using. Here I shall be using 'content' in a way closely related to the first and broadest sense discussed, as the features of an attitude other than the kind of attitude it is. But as we shall see, the distinctions that Kaplan and Fodor make are important.

Any view about the content individuation must sustain the capacity of content to function in the psychological explanation of individual subjects' inferences and behavior. Individuation of content must be fine-grained enough to account for differences in inference and behavior, but it must also be coarse-grained enough to capture salient similarities in the production of inference and behavior from subject to subject, and for a single subject across time.⁷ Frege's test is strongly evidential for fineness of grain in the intrapersonal case.⁸ 'Hesperus' and 'Phosphorus' refer to the same entity, but if the identity statement Hesperus = Phosphorus would be informative for a subject, then an acceptable psychological theory must make a distinction between the concepts 'Hesperus' and 'Phosphorus' for that subject (supposing she has both concepts), which is to say that in her case it must make a distinction between the contributions to content (i.e. the cognitive value) of these concepts. In general, if a subject has the concepts 'A' and 'B,' and if 'A = B' would be informative for her, then 'A' and 'B' are different concepts for that subject, and thus for her they differ in cognitive value.

Let us recall why Frege's test provides strong evidence for differences in contribution to content in the intrapersonal case. Content plays the pivotal role in the psychological explanation of inference and behavior. If Hesperus = Phosphorus would be an informative identity for a subject, then for that subject beliefs and desires involving the concept 'Hesperus' can make significantly different contributions to explanations of inference and behavior from those involving 'Phosphorus.' If Ariane does not know that Hesperus = Phosphorus, her belief that Hesperus is a planet cannot be regarded as having the same content as the belief that Phosphorus is a planet, because she might infer that Hesperus revolves around a star from her belief that Hesperus is a planet, and yet fail to infer from this same belief that Phosphorus revolves around a star. And her desire to visit Hesperus just once cannot be counted the same as her desire to visit Phosphorus

just once, because supposing her desire to visit Hesperus just once has caused her to make arrangements to land on Hesperus, this would not preclude her from making further plans to land on Phosphorus. Thus a psychological theory that aims to track significantly different contributions to explanations of inference and behavior must distinguish Ariane's concept 'Hesperus' from her concept 'Phosphorus,' despite the fact that these concepts are coreferential.

One might worry that Frege's test is inadequate because it threatens to distinguish cognitive value wherever there is any difference in functional role, and that thus this test will also undermine the psychological generalizations we want to make. But this worry can be set aside. Although, for example, the functional roles of my concepts 'fourteen days' and 'fortnight' are different, Frege's test does not distinguish them. My thoughts involving 'fortnight,' are more likely to make me think 'what a quaint thought!' and to call to mind Burge's article "Belief and Synonymy" than are my thoughts involving 'fourteen days.' But nevertheless, a fortnight = fourteen days is not informative for me. Hence, there are differences in functional role where Frege's test will not distinguish cognitive value.

Frege's test poses a problem for some externalist theories of content. Consider, for example, the Millian account of proper names discussed by Saul Kripke, according to which coreferential proper names are intersubstitutable in belief contexts salva veritate, and make the same contribution to semantic content.⁹ In an extension of this semantic view to the explanation of inference and behavior, coreferential proper names would make the same contribution to psychological content. Kripke's puzzle about belief provides a persistent difficulty for such an account, and also, he argues, for any theory of belief attribution.¹⁰ Pierre comes to have a belief which he expresses as "Londres est jolie." By the rules of translation and disquotation, we English-speakers attribute to him the belief that London is pretty. But Pierre lands up in London, in a bad part of town, and he acquires a belief that he expresses as "London is not pretty," which, by the rule of disquotation (and perhaps translation from Pierre's idiolect) we attribute to him as the belief that London is not pretty. He all the while maintains the belief 'Londres est jolie,' and

thus we would continue to attribute to him the belief that London is pretty. Accordingly, we attribute contradictory beliefs to him, but this would seem to involve a mistake. The anti-Millian point of this case is that attributing a contradiction in belief to Pierre seems unavoidable if coreferential proper names make the same contribution to content. But Kripke argues that the puzzle raises a problem not just for a Millian view, but for all theories of belief attribution, because they all would endorse the rules of translation and disquotation, and it is not obvious that any other conceivably controversial elements are required to construct the puzzle.

As Loar has shown, the anti-Millian point of Kripke's puzzle can be expressed without using the notion of contradiction. Even if we conceive of Pierre landing up in an upscale part of London, and forming the belief that he expresses as "London is pretty," we could not count this new belief to be the same as the belief he acquired in France, since they are not inferentially identical. It seems to follow that coreferential proper names must sometimes make different contributions to content.¹¹ And as Akeel Bilgrami argues, Loar's argument allows us to see that the anti-Millian point of Kripke's case can be expressed by a Fregean puzzle. For Pierre Londres = London would be informative, and thus for him, these proper names differ in cognitive value.¹² Hence, Frege's test provides strong evidence that for Pierre 'Londres' and 'London' make distinct contributions to content.

Another example of Kripke's reveals that an informative identity can be constructed even when the coreferential proper names are the same (when the vehicles, in Fodor's sense, are the same). A subject may not realize that Paderewski the statesman is identical to Paderewski the pianist, and therefore the identity Paderewski = Paderewski may be informative for her. In addition, when the Millian view is applied to demonstratives, coreferential demonstrative expressions make the same contribution to content. Burge points out, however, that Fregean puzzles can be constructed not only for coreferential and identical proper names, but also for coreferential and identical demonstrative expressions.¹³ This = this can be informative if the first 'this' is accompanied by a demonstration of the svelte Orson Welles in The Lady from Shanghai

and the second by a demonstration of a more corpulent Welles in Touch of Evil.

Accordingly, in each of these cases, we have strong evidence that the two expressions flanking the identity sign indicate different contributions to content. As the "Paderewski" example shows, the divergence in cognitive value between the two expressions cannot always be explained by a difference in vehicles. Moreover, even when vehicles are distinct, it is dubious that the distinction between vehicles always or even typically accounts for differences in cognitive value. Plausibly, a further contribution is sometimes provided by the beliefs associated with the expressions. For example, in Kripke's case, the beliefs Pierre associates with 'Londres' might well be different from the beliefs he associates with 'London.' But even if we make this move, Kripke would argue that the problem for belief attribution remains unsolved. We can imagine that many French people have just the information about London that Pierre acquired in France, and that counterfactually, they would be in Pierre's predicament. Yet we would want to make unembellished attributions, like 'Madeleine believes that London is pretty,' to them. So if we were to apply our standards for belief attribution consistently, we would still want to attribute to Pierre the belief that London is pretty, and this is what generates Kripke's puzzle.

A common objection to Burge's view about content individuation is that it sometimes renders contents equivalent when they must explain significantly different patterns of inference and behavior, and at other times distinguishes contents that must explain very similar patterns of inference and behavior. To illustrate, let us examine a case Burge has used to argue for his view.¹⁴ Alfred is a member of the English-speaking linguistic and social community, and he correctly applies the term 'arthritis' in many situations. But he has recently developed a pain in his thigh, and he has come to believe "I have arthritis in my thigh," even though in his community arthritis is classified by the experts as a disease solely of the joints. Now imagine a counterpart to Alfred, who is identical to him in every non-intentional individualistic feature, but in his community the word 'arthritis' refers to certain rheumatoid ailments, including arthritis, but also including certain ailments of the muscles and tendons, and in this community his belief (expressed

in his language) "I have arthritis in my thigh" is true. We English-speakers would attribute a belief involving our concept of arthritis (with our word 'arthritis' in oblique position) to Alfred, but not to his counterpart, even though there are no non-intentional individualistic differences between them. To attribute a belief in our language to the counterpart, we would have to devise a translation, such as 'tharthritis,' for his term 'arthritis.' Hence, the content of Alfred's attitude, and more specifically, his concept of arthritis, is dependent not simply on non-intentional individualistic facts about him, but also on facts about his social environment.

For the thought-experiment to support Burge's view, it must be plausible to attribute contents including the concept 'arthritis' to Alfred, just as to an expert on earth, and to deny the concept 'arthritis' of Alfred's counterpart. But many report the individualist intuition that the behavior and inferences of Alfred and his counterpart are explained by the same psychological states. Furthermore, although the differences in behavior between Alfred and the expert might sometimes be explained by the different 'arthritis' beliefs they hold, a significant enough divergence in such beliefs could conceivably force us to attribute different concepts to them. For the view that concepts, when attributed in psychological explanations, are completely independent of surrounding beliefs is implausible.¹⁵

Burge and others have developed theories that disavow such individualist intuitions. My aim is to develop a view that preserves these intuitions while explaining the considerations that pull in the opposite direction, and in addition supplies a solution to Kripke's puzzle, and provides a method of content individuation that can sustain the psychological generalizations we want to make. Such a view will not obviously undermine an externalist position of the sort that Burge has advocated. But it will provide an alternative to such a position, and its plausibility will derive from the force of the individualist intuitions, and from the attractive ways in which it can meet the other desiderata of a theory of content.

Both Kripke's puzzle and Burge's thought-experiments depend crucially on taking ordinary attribution practice to be strongly indicative, if not definitive, of the nature of the content of an attitude. Hence, in developing an alternative to their externalist theories, one might attempt to undermine this confidence in ordinary attribution practice. One might begin by differentiating two perspectives from which attributions can be evaluated. Burge himself, in a discussion of Frege's views on the sense of beliefs involving proper names and indexicals, distinguishes the public from the psychological functions of attribution. What is "communicated through ordinary language when proper names, or other indexicals are used" is distinct from what is required "to solve the problems of cognitive value and oblique belief contexts."¹⁶ John Perry, Donald Davidson, and Michael Devitt have also suggested a distinction of this sort, and Davidson and Devitt have argued that it can be brought to bear on Burge's thought-experiments.¹⁷ Devitt argues that in attribution, the purpose of psychological explanation can be distinguished from the aim of learning about the world:

Folk ascribe thoughts not only for psychological purposes but also to learn about the world. This dual purpose raises the possibility that the folk do not regard all the content they ascribe as relevant to the explanation of behavior.¹⁸

I shall attempt to develop this type of suggestion in greater detail.

Belief attributions have various purposes. One goal of belief attribution is to communicate to others evidence about features of the world represented by a belief. That this is an important function of belief attribution is clear. People's beliefs are a significant source of evidence about what they represent, and public attribution of beliefs is an important means for communicating this evidence. If, however, communication of evidence about the world is one's aim in attributing a belief to a subject, the exact way the subject represents that evidence might well not be pertinent. For example, even if a subject cannot distinguish elms from most other deciduous trees, and even if she believes that elms and beeches are the same type of tree, the attribution of her belief as "She believes that elms are subject to epidemics" can still communicate evidence about elms. Despite

the subject's meager conceptual ability, if her beliefs have certain contextual relations to elms, perhaps involving a causal chain, attributions of her beliefs as being about elms can still provide evidence about elms to an audience.

Belief attribution has another fundamental purpose, the part it plays in the psychological explanation of inference and behavior. And when this is the main goal, attribution will tend to have different emphases than it does when the aim of communication of evidence about the world dominates. When communication of evidence about the world is the chief aim, carefully delineating all aspects of a subject's attitudes may be irrelevant. But specifying the nature of a subject's attitudes very precisely may be critical for psychological explanation, because the exact nature of the attitude can make a difference for her inferences and behavior.

When communication of evidence about the world is at issue, belief attribution can prescind from the details of the ways in which entities are represented. De re belief attributions often fit this description. When I make the de re attribution of Pierre's belief, 'London is such that Pierre believes it not to be pretty,' the mode by which Pierre represents London might be of no concern to me because in attributing this belief I am interested mainly in communicating evidence about London. And that the mode by which Pierre represents London is of no concern to me may be indicated by the fact that in my attribution coreferential expressions are substitutable for 'London' salva veritate. Moreover, although in de dicto attributions delineating the subject's mode of representation is always an aim, communication of evidence about the world might function as an additional purpose there as well, and might have an effect on the character of the attribution. My attribution 'Pierre believes that the capital of England is pretty' may come close to specifying Pierre's actual mode of representation of London, while 'Pierre believes that the largest city in England is pretty' does not, and, accordingly, substitution of coreferential expressions salva veritate in my attribution fails. But nevertheless the precedence taken by the goal of communication of evidence about the world might render my specification of further details about the way Pierre represents London unnecessary.

Let us now reexamine Kripke's puzzle with these reflections in mind. The attribution to Pierre of the belief that London is pretty can communicate evidence that London is pretty. And further, the attribution to him of the belief that London is not pretty can communicate evidence that London is not pretty. But although these attributions might communicate conflicting evidence, and thus provide evidence for incompatible states of affairs, we have yet no reason to believe that there is a contradiction in the way Pierre represents London. For if Pierre's psychological states were more accurately specified -- described more precisely than is required for the aim of communication of evidence about the world -- any such contradiction might well disappear.

Furthermore, distinct tasks in psychological explanation may demand different levels of precision in the specification of psychological states. If we want to explain why Pierre, while still in France, says to the travel agent "Je voudrais voyager à Londres," we can do so partly by attributing to him the belief that London is pretty. If we want to explain why Pierre, now in England, chooses to get over his bad mood by listening to music rather than by taking a walk, we can do so partly by attributing to him the belief that London is not pretty, or in French, que Londres n'est pas jolie. Given our typical explanatory interests in such contexts, we need not qualify our attribution of the concept 'London' or of the concept 'Londres' to Pierre, and indeed so far -- and this is often the case -- psychological explanation does not demand greater precision in the specification of content than does communication of evidence about the world. But when we must explain how he can believe both 'Londres est jolie' and 'London is not pretty,' our characterizations of Pierre's attitudes must become more detailed and precise. In particular, more must be said about the way in which he is thinking about London -- his mode of presentation of London. One way to accomplish this is to hedge our attribution of the concept 'London' by specifying that he has not realized that Londres = London, and therefore Pierre's beliefs involving his use of 'Londres' are inferentially isolated from those involving his use of 'London.'

Throughout his article, Kripke stresses the point that there is some terminology that

suffices to explain Pierre's situation. He argues, however, that this does not solve his puzzle.

But beware of one source of confusion. It is no solution in itself to observe that some other terminology, which evades the question whether Pierre believes London is pretty, may be sufficient to state all the relevant facts.... But none of this answers the original question. Does Pierre, or does he not, believe that London is pretty? I know of no answer to this question that seems satisfactory.¹⁹

As we have seen, our attribution of beliefs to Pierre can communicate evidence that London is pretty and evidence that London is not pretty. We can legitimately make such attributions, given the understanding that they reflect not the precise nature of Pierre's psychological states, but the evidence about the world that his beliefs can provide, as well as the requirements of many cases of psychological explanation. But in Kripke's case, psychological explanation requires a more precise specification of how Pierre represents London, and thus we would not attribute his French belief simply as "Pierre believes that London is pretty." Rather, we would stress the details of his mode of presentation by hedging any less precise attributions. As a result, there will not be an unqualified yes/no answer to Kripke's question. But this raises no intractable problem for belief attribution. Just as for answers to "Was French foreign policy responsible for the geopolitical decline of Spain in the seventeenth century?" and "Is the ocean blue?" an answer to Kripke's question must be qualified. Reality is too complex for simple yes/no responses to such questions to be satisfactory. The content of Pierre's attitude is too intricate, perhaps both in its conceptual and indexical features, for Kripke's question to have an unqualified yes/no answer.²⁰

We can also use these observations to explore an alternative analysis of the issues raised by Burge's example. Burge provides good reason to believe that ordinary belief attribution is often non-individualistic in the sense that our ordinary attributions of content are dependent on facts about a subject's physical and social environment. From this fact about ordinary attributions he concludes that belief contents themselves have externalist individuation-conditions. Let us, however, attempt to develop an account of content that takes seriously the individualist intuitions

that psychological explanation could require the cognitive value of Alfred's and his counterpart's concepts to be the same, and the cognitive value of Alfred's and the expert's concepts to differ. Rather than concede an externalist position about content individuation, we might then instead regard these intuitions as providing evidence against taking ordinary attribution as determinative of cognitive value. Throughout his writings, Burge assumes that ordinary attribution practice yields a notion of content adequate for psychological explanation.²¹ But while ordinary attribution practice may be adequate to psychological explanation in typical cases, it might be that in more challenging cases, those that more accurately disclose the real nature of attitude contents, ordinary practice is inadequate.

Let us reexamine Burge's thought-experiment to see whether the intuitions that oppose his view can be accommodated, while keeping in mind the distinction between attribution aimed at communication of evidence about the world and attribution geared to psychological explanation. Alfred and his counterpart are identical with respect to all non-intentional, individualistic characteristics. Alfred's English-speaking society uses 'arthritis' to refer to a rheumatoid ailment of the joints, while in the counterpart's society this term is used to refer not only to this ailment of the joints, but also to ailments of the muscles and tendons. Burge argues convincingly that we would ordinarily attribute beliefs using our term 'arthritis' -- without qualification -- not only to the expert, but to Alfred as well. But this fact might be explained by our interest in communicating, by means of belief attribution, evidence about the features of the world represented by a belief, and by the requirements of psychological explanation in many ordinary contexts. First, the goal of communication of evidence about the world through attribution requires that we use terms that have public linguistic meanings, and that we use these terms in accordance with their public linguistic meanings. And thus this goal gives us reason to make attributions of Alfred employing the term 'arthritis' with its public linguistic meaning, and not using 'arthritis' while intending a linguistic meaning other than the public one, or using a made-up term not in the public language -- even if the term 'arthritis' fails to capture the nature of his

concept accurately. Further, even though Alfred's knowledge of arthritis is too limited to preclude even fairly elementary errors -- measured against the standard of public linguistic meaning -- the combination of his conceptual and contextual relations to arthritis might well be sufficient for attribution of his beliefs as being about arthritis to communicate evidence regarding the rheumatoid ailment of the joints. Given Alfred's conceptual and contextual relations to arthritis, therefore, attributions using the term 'arthritis,' unhedged, with its public linguistic meaning, might well be sufficient to satisfy the aim of communication of evidence about the world, even if psychological explanation could require that we not attribute to him the same concept as the expert. Moreover, attribution to Alfred using our term 'arthritis,' unhedged, would likely be sufficient for many cases of psychological explanation. But if we want to specify the nature of his concept more precisely -- and perhaps we must do so in order to explain some of the differences between the inferences and behavior of Alfred and the expert -- attributions of our concept 'arthritis' to Alfred might well have to be qualified.

Our predisposition to deny attribution of our concept 'arthritis' to Alfred's counterpart might also be explained by the aim of communication of evidence about the world. In his thought-experiment, Burge specifies that the public linguistic meaning of 'arthritis' in the counterpart's community is not the same as the linguistic meaning of 'arthritis' in our community. Hence, to attribute a belief using our term 'arthritis' to Alfred's counterpart would be to attribute to him a term intending a linguistic meaning foreign to his community. Most often communication of evidence about the world by attribution is directed towards the subject's own community. Typically, therefore, attribution using the linguistic meanings of the subject's community facilitates communication of evidence about the world, while such communication is frustrated by attribution using linguistic meanings foreign to the subject's community. Consequently, our attribution practice is strongly disinclined to attribute to a subject terms intending linguistic meanings alien to his community, and we are accordingly averse to using 'arthritis' with our linguistic meaning in making attributions to Alfred's counterpart. Nevertheless,

if our aim is psychological explanation, our attributions, together with our hedgings and qualifications, may yet be forced to exhibit a convergence of the concepts of Alfred and his counterpart.

Many philosophers, such as Loar, Block, and Fodor, have argued that variation in standards of attribution indicate that there are two types of content -- wide, externalist content, and the narrow content required for psychological explanation of inference and behavior. In the view I am developing here, as in Bilgrami's, content is unitary, and its true character is most accurately approximated in the context of psychological explanation. The arguments of Burge and others do not, all by themselves, show that "wide content" qualifies as a separate kind of real psychological content. "Wide content," it would seem, is derived from attributions that in part aim to facilitate communication of evidence about the world. But this purpose might be viewed as having a tendency to deflect the aim at truth about the attitudes of the subject in the interest of communicating evidence about the features of the world that the attitude represents. Thus when our interest is in communication of evidence about the world, we attribute to Pierre the belief that London is pretty, disregarding the more complex nature of his representation of London. In this case, we, as attributors, might regard ourselves as not engaged solely in specifying what is in Pierre's mind -- in describing the real nature of his attitudes -- but also by the additional and potentially conflicting goal of communicating evidence about the features of the world that Pierre's beliefs are about. When, by contrast, our interest is psychological explanation, our focus is solely on Pierre's psychology, and on representing Pierre's attitudes accurately, or as accurately as the particular task in psychological explanation requires. In this picture, therefore, attribution of "wide content" incorporates a tendency to represent the content of a subject's attitudes imprecisely, whereas when our interest is psychological explanation, attribution is seen as approaching content as it really is.

Correlatively, in this view it is not the case that there are two different types of truth-conditions for content attribution. When one aims at communication of evidence about the world,

there may be good reasons for attributing a "content" that diverges from the real nature of the attitude. Such divergence may have genuine pragmatic value, but it does not aim solely at truth about the content of the attitude. And thus success in the pursuit of communication of evidence about the world might quite readily conflict with the satisfaction of the truth-conditions of content attribution. Since pursuit of the goal of communication of evidence about the world has a tendency to motivate an imprecise characterization of the contents of subjects' attitudes, it would be a mistake to posit an additional sort of truth-condition that corresponds to this purpose. When, by contrast, one's quest in attribution is psychological explanation, one aims more directly and less equivocally at the truth about subject's attitudes, and therefore also at the satisfaction of the truth-conditions of attribution.

Furthermore, if ordinary content attribution serves partially to communicate evidence about the world, and this purpose tends to conflict with representing content as it really is, then concepts as they are employed in ordinary content attribution do not have externalist individuation-conditions of the sort we have examined. For if the factors in thought-experiments that generate externalist conclusions can be explained as a function of the goal of communication of evidence about the world, and if this purpose deflects the aim at truth about attitude contents, then we should not look to attributions part of whose purpose is communication of evidence about the world for the individuation-conditions of concepts. These individuation-conditions are not best revealed by instances of pragmatically valuable attribution practice which has a tendency to generate false attributions, but rather by attributions that aim at truth about attitude contents.

In summary, our objective was to develop a view that solves Kripke's puzzle and explains the results of Burge's thought-experiments while preserving the key individualist intuitions. In the picture we have constructed, attitude attributions have (at least) two aims, communication of evidence about the world and psychological explanation. Ordinary attribution typically combines both purposes. But ordinary attribution need not be regarded as unequivocally tending to accuracy about the nature of attitudes. For while the goal of psychological explanation inclines to

subservient truth in attribution, the purpose of communication of evidence about the world has a tendency to deflect the aim at truth about attitudes. In addition, psychological explanation typically does not require maximal precision in specification of attitudes. Thus the ordinary attributions of Kripke's puzzle and Burge's thought-experiments need not be viewed as representing attitudes accurately. Consequently, there might well be room for the concepts of Alfred and his counterpart being the same, and perhaps also for the concepts of Alfred and the expert being different, and Pierre's beliefs might really not be puzzling after all.

III

Our account must now provide a method for content individuation that can support the psychological generalizations we want to make. Let us first consider the causal theories, developed by Dretske, Millikan, and Fodor, which supply a method for content individuation that has become prominent in recent years.²² The common element in these causal theories is that the cognitive value or psychological meaning of representations is provided by entities or properties that cause them. Causal theories differ in their ways of handling what is perhaps their deepest difficulty, the disjunction problem, which has been outlined by Fodor.²³ To illustrate, 'cow' representations are caused not only by cows but also by bushes on dark nights, and thus a simple causal theory would have 'cow' mean not just cow but rather cow or bush-on-a-dark-night. The disjunction problem forces causal theories to provide a way of selecting the right causes from all the possible causes of a representation. In Millikan's theory, the right causes are singled out by the notion of proper biological function. In one statement of this view, the content of a belief is C just in case the role that beliefs of this type play, when the mechanisms that produce this belief are fulfilling their proper functions as determined by evolutionary biology, involves being caused by C's.²⁴ In Fodor's asymmetrical dependence theory, 'cow' means cow and not cow or bush-on-a-dark-night because 'cow' tokens are caused by things in virtue of being cows, and the existence of 'cow' tokens caused by things in virtue of their being bushes on dark nights depends on the

existence of 'cow' tokens caused by things in virtue of their being cows, and not vice-versa. Fodor's idea is that there is a hierarchy of causal dependence among causes of tokens of a mental representation-type, and the property that specifies the psychological meaning or cognitive value of that representation is causally most fundamental.

Many difficulties have been raised for each of these theories.²⁵ A problem from the perspective of someone swayed by individualist intuitions is that causal theories would seem to acquiesce to Burge's analysis of his thought-experiments. For example, by a causal theory, the "arthritis" concepts of Alfred and his counterpart would have to differ, since their external-world causes differ -- the disease called "arthritis" on earth is not the same disease as the disease called "arthritis" on twin earth. In addition, the 'arthritis' concepts of Alfred and the expert would seem to be identical, no matter how different the associated beliefs were, as long as these concepts were caused by the same disease. Causal theories, therefore, seem unable to preserve the individualist intuitions.

A further problem for causal theories would concern anyone, whether or not her intuitions are individualist. One of the major difficulties for causal views is that they fail to accommodate the results of Fregean tests. Loar's revised version of Kripke's puzzle about belief is especially illustrative here. Pierre acquires the belief 'Londres est jolie' in France, and then lands up in a pretty part of London, where he acquires the belief 'London is pretty,' without realizing that Londres = London. By Frege's test, his concept 'Londres' is distinct from his concept 'London.' Psychology must hold them distinct, because they play significantly different roles in Pierre's inferences and in the production of his behavior. But their causes might be precisely the same -- indeed, the views of London that caused the 'Londres' concept might be the very same as those that caused the 'London' concept.²⁶

Fodor proposes to handle this objection by saying that although the belief contents of Pierre's beliefs 'Londres est jolie' and 'London is pretty' are the same, nonetheless the beliefs differ.²⁷ Hence Fodor hopes to retain a causal theory of content, while eschewing a causal theory

of belief. As we have seen, Fodor's view is that a belief is a function of four factors, type of attitude, content, vehicle, and functional role. According to his analysis of Kripke's puzzle, the contents and type of attitude of Pierre's two beliefs are the same, whereas their vehicles, or the functional roles of their vehicles, differ.

Such a solution would be satisfactory, if, among other things, Fodor provided us with an account of the individuation of vehicles and their functional roles that would sustain plausible generalizations in psychological explanation. But surprisingly, Fodor argues that psychological explanation can dispense with vehicle and functional role, and embrace only content in his own sense and, presumably, type of attitude.²⁸ Fodor hopes to exclude vehicles and functional roles from psychological explanation precisely in order to avoid their messy individuation-conditions, while he retains causally individuated content and type of attitude because their individuation-conditions are less problematic. Limited by the resources of type of attitude and content in Fodor's sense, however, psychology will be unable to explain the differences in inference and behavior that emerge from Pierre's failure to identify London and Londres. By Fodor's characterization of content, the contents of Pierre's beliefs 'Londres est jolie' and 'London is pretty' are the same, and they therefore play the same role in psychological explanation. Fodor does have elements in his theory that could explain the differences in inference and behavior resulting from Pierre's failure to identify London and Londres. But since he lacks a way to specify when these elements are the same and when they are different, his theory fails to solve one of the fundamental problems for accounts of psychological explanation.

One of the main reasons why theorists have been drawn to causal views about content is to provide a theory that meets the requirements of naturalism, and causal theories are regarded as sufficiently naturalistic in virtue of their capacity to incorporate reductionism. But naturalist requirements, I believe, can be met without any type of reductionism. In my view, furthermore, reductionist theories of content should be rejected for the reason that they cannot individuate content with the accuracy required for psychological explanation. While I will not argue for this

general claim here, it is well-illustrated by one of Fodor's counterexamples to Millikan's biological theory of the content of desires.

According to Millikan's view, the content of a desire is provided by its proper biological function. In one formulation of her account, the content of a desire is C just in case the role that desires of this type play, when the mechanisms that produce this desire are fulfilling their proper functions as determined by evolutionary biology, is to cause the organism to secure C's.²⁹ Fodor's counterargument examines our desire for sweets. By evolutionary biology's account of the proper function of the mechanisms that produce the desire for sweets, the role of this desire is to get the organism to ingest sugars. Consequently, on Millikan's theory, the desire for sweets is correctly described as a desire for sugars. But the desire for sweets is satisfied by saccharine and Nutra-Sweet, and thus, the content of the desire for sweets is just what it seems to be, for sweets, and not for sugars. The content of the desire for sweets, therefore, cannot be derived from its proper biological function.³⁰

Millikan might respond that the desire for sweets might really be the desire for sugars despite Fodor's counterexample, because the possibility of a desire for A being satisfied by something that only appears to be A provides no reason to believe that the desire is really for something that appears to be A. The desire for something sweet being satisfied by Diet Coke (which contains Nutra-Sweet) is much like a desire for a Vermeer painting being satisfied by a van Meegeren forgery. The desire for the Vermeer being satisfied by the forgery does not mean that the desire for a Vermeer is really just a desire for a painting that looks like a Vermeer. Hence, the possibility of our system's being fooled by Nutra-Sweet is not adequate reason to reject the biological account of desire content.

One odd consequence of this response is that despite appearances, Diet Coke really fails to satisfy the desire for sweets, just as the forgery really doesn't satisfy the desire for a Vermeer. This is counterintuitive, perhaps because if one were asked if one's desire for sweets was satisfied after it was pointed out that the Diet Coke one just consumed contained no sugars, one might

well say "yes," whereas if one were asked if one's desire for a Vermeer was satisfied after it was pointed out that the painting one just acquired is a forgery, one would say "no." This provides reason to believe that Fodor's "desire for sweets" case is not one of being fooled by appearances.

But a further counterexample might prove more convincing. By evolutionary biology's account of the proper functions of the mechanisms that produce the desire for sex, the role of this desire is to get the subject to reproduce. Thus on Millikan's account, desire for sex must actually be the desire for reproduction. An unintuitive consequence of this view is that desire for sex would not really be satisfied unless reproduction were attained. One might hold out for the claim that desire for sex is only apparently satisfied when reproduction is not achieved, but this view is far from plausible.

At the root of the difficulty for Millikan's view is that desire for sex plays a role in psychological explanation that is systematically different from the role played by desire for reproduction, despite the fact that the evolutionary purpose of the desire for sex is to get the organism to reproduce. This problem, one might note, is not generated by assuming intrinsic intentionality, or supposing a privileged introspective point of view, but solely by reflection on the actual practice of psychological explanation. And thus the relation between evolutionary biology and psychology is not as close as Millikan makes it out to be. Evolution can achieve its aims by granting us a system of beliefs and desires -- a psychology. But this psychology is significantly independent of evolutionary theory for the individuation of its kinds. In the case of sexual desire, we form a desire not for reproduction, but for the intimacy and sensory pleasures associated with sex. This strategy is perhaps a better way for evolution to secure reproduction than by having us form a desire to reproduce per se. Similarly, given our natural environment, a desire for the sensory pleasure provided by sweet tastes may well be a more efficient way to make us ingest sugars than would a desire for sugars per se, especially if this alternative would involve representing their chemical structure. In both cases, psychological content cannot be determined by evolutionary biological considerations alone.

These reflections suggest that it is advantageous to regard individuation of psychological content as largely independent of evolutionary biological considerations. A reductionist evolutionary biological theory seems unable to individuate psychological states with the accuracy required for psychological explanation. More generally, I favor the nonreductive view according to which content individuation is largely independent of all more basic scientific theories.³¹ In recent papers, Stephen Stich and Michael Tye have argued convincingly that the nonreductive strategy can be applied to a theory of content.³² They maintain, correctly, I believe, that for content to be naturalistic, it need not be reduced to the non-intentional, for if psychology meets the criteria of a true scientia in its own right, and the entities it posits are constituted by entities over which physics quantifies, worries about its naturalistic credentials can be laid to rest.

IV

If one wishes to reject the various well-known externalisms, such as the views of Burge and Millikan, and if one wants to provide adequate foundations for psychology, one must provide reason to believe that a satisfying alternative theory of psychological content individuation is in the offing. Externalist theories provide a workable, although perhaps imperfect theory of content individuation, and if one has no indication of a better option, it would be best to appropriate some form of externalism despite its counterintuitive features. Psychology is very likely to require some method of content individuation, and it might have to settle for the best one available despite imperfections. I shall argue that there is a promising alternative to the foremost externalist theories, and that this alternative that does not suffer from the difficulties faced by the functional role theories that have been advanced.

According to functional role theories, the content of a mental state is a function of the perceptual inputs, behavioral outputs, and the other mental states to which the state is causally related. The main problem for this type of view is that it appears to lack any principled and satisfying method for classifying token-distinct psychological states as having the same content.

My belief that it is raining likely has a somewhat different causal role than your belief that it is raining, and so functional role theory might well class them as being beliefs of different types. But this method of individuation precludes the kinds of generalizations in psychological explanation that we want to make.

A strategy for content individuation within a framework that rejects certain well-known externalisms has recently been developed by Bilgrami. According to his proposal, the cognitive value of a subject's concept is individuated solely by the beliefs associated with that concept.³³ His strategy, as he himself points out, must have the resources to solve three problems.³⁴ The first of these we have already explored: A theory in which the cognitive value of a concept is individuated solely by the beliefs associated with the concept (at a time) will cut cognitive value too finely, since when cognitive value is determined in this way it will not sustain plausible psychological generalizations. Second, if all of a subject's associated beliefs were relevant to individuating the cognitive value of a concept, any change in belief would amount to a change in the cognitive value of a concept. And if any change in belief would amount to a change in cognitive value, then a subject could not deepen or broaden her grasp of any one concept. Third, if the cognitive value of a concept were determined by associated beliefs, then concepts that these beliefs involve would constitute the psychological meaning (although perhaps not the linguistic meaning) of that concept, and hence, they would be analytic of it (in one traditional sense).³⁵ A theory of this sort, therefore, would seem to reintroduce a version of the analytic/synthetic distinction, which, for Quinean reasons, might be unappealing.

Bilgrami proposes to solve these problems by an appeal to context. Only in some abstract sense is the cognitive value for some term individuated by all of the subject's beliefs associated with it. In any specific situation, the context will select a subset of those beliefs to determine the cognitive value for the term. Alfred and the expert may share the belief that Alfred's mother has arthritis in her fingers, in part because in the context of attribution all of the beliefs relevant to individuating the cognitive value for 'arthritis' are possessed by both subjects. In addition, change

of belief is possible without change of cognitive value because the beliefs that a context designates as relevant to individuating the cognitive value for a term may not include those that change.³⁶ Finally, since distinct beliefs are selected to provide the cognitive value or psychological meaning for a term in different contexts, it could well be that no particular concepts provide the psychological meaning for a term in all contexts, and thus no particular concepts would be analytic of a term.³⁷

Bilgrami does not tell us in any detail, however, how context might play the role he envisions, for example, how context indicates which beliefs of Alfred and the doctor would be relevant to individuating the cognitive value for the term 'arthritis.' The details of his contextual theory are important, because they would constitute the principles that determine how generalizations in psychology are to be made, and how cognitive value can remain stable despite variation in associated beliefs. The seriousness of the problem becomes clear when we bring to mind that a theory of attribution, at minimum, must meet two requirements. First, it must illuminate, rather than merely reduce to, our intuitions about attribution in particular cases. Second, the theory must be testable against such particular intuitions about attribution. If, however, Bilgrami fails to provide principles that specify the relations between contexts and attributions, the theory does indeed threaten to reduce to our intuitions about particular cases. If he insists that nevertheless the theory does more work than such intuitions do, but he supplies no principles that specify the relations between contexts and attributions, we will have no clear way of distinguishing between the deliverances of the theory and our intuitions. And if we cannot make these distinctions, we will have no way of testing whether the theory is true.

It is doubtful, however, that the principles Bilgrami needs to supplement his contextual theory can be provided. Perhaps the relations between contexts and attributions are much too complex to be formulated in a useful way. One might thus argue that a contextual account without general principles for individuation of content might be the best theory available. But one should agree that if there were a theory that supplied such principles, then there would be some

reason to favor such a view.

V

Practices in other sciences suggest a more promising strategy for individuating content. When we classify entities as members of the same or different kinds in chemistry and biology, we often look to structural features of these entities. The stuff in streams and the stuff in oceans is of the same type because it has the same chemical structure. We consider different human beings to be members of the same species partly because of similar genetic structure. Recent work in cognitive psychology indicates that many concepts also have structure of various sorts, and consequently there might be a basis, similar to those in other sciences, for evaluating questions about sameness and difference of these concepts, and also of the content of which these concepts are components. In the last two decades, significant insight into conceptual structure has been achieved. But much more remains to be done, and accordingly, what follows is not meant to constitute an actual theory of content individuation, but rather a suggestion for the direction such a theory might take.

Conceptual structure will be valuable for individuating concepts and content if from among the beliefs associated with a concept it provides a way of selecting some as more weighty than others. Since concepts are essentially connected with a complex of abilities that includes classification, identification, and recognition, one might aim to isolate the associated beliefs that are most salient for these abilities.

Cognitive psychologists have recently argued for what has come to be known as psychological essentialism.³⁸ This is not the view that things have essences, but rather that our representations of things of some kinds are structured by beliefs of them as having essences. According to this theory, these experimentally identifiable beliefs play a much more important part than others in our classification of such things. For example, many natural kind and some artifact concepts are structured by an assumption that internal properties of the things falling under those

concepts are causally responsible for their surface properties, and that these internal properties are more important than their surface properties for the things being what they are. Hence, when we classify such things, their internal properties have a much more influential function in carrying out these tasks than do their surface properties. Sometimes actual concepts of the internal properties play the crucial part in the structure of the concept, but often a placeholder conception, like 'whatever the experts believe to cause the symptoms of the disease called "arthritis"' or 'whatever internal properties really cause the symptoms of the disease called "arthritis"' has the decisive role.³⁹

One of the most significant features of psychological essentialism is that it provides a principled and empirically testable way to determine which beliefs play the central role in structuring certain concepts. Let us consider two examples. In one experiment performed by Susan Gelman and Henry Wellman, four- and five-year olds were asked for items like dogs and cars whether their insides or their outsides were more important for being what they are. For example, they were asked to consider the following situations: "What if you take out the stuff inside a dog, the blood and bones and things like that and got rid of it and all you have left are the outsides?" and "What if you take the stuff outside of the dog, you know, the fur and got rid of it and all you have left are the insides?" They were asked two kinds of questions about the transformed entity, an identity question like "Is it still a dog?" and a function question such as "Can it still bark and eat dog food?" For the identity question, 72% of the responses indicated that the entity is no longer what it was when the insides are removed, while 35% indicated that it is no longer what it was when the outsides are removed. 92% indicated that the functions are lost when the insides are removed, 29% when the outsides are removed.⁴⁰ Experiments of this sort provide evidence that young children's conceptions of internal properties of dogs and cars -- despite their lack of detailed knowledge of these properties -- play a more crucial role in classifying such things than do conceptions of surface properties.

In another study, Gelman and Wellman told four-year olds about an animal that was

brought up in an environment suited to a different species, and questioned the children in order to determine whether they believed the animal developed innate potentialities, or instead displayed the properties associated with the environment of upbringing. For example, they asked about a cow who was raised by pigs what her mature physical features would be like, for example whether her tail would be straight or curly, and what her mature behavior would be like, for instance whether she would say "moo" or "oink." Overall, children answered on the basis of innate potential 85% of the time. Experiments like this provide evidence that young children's conceptions of internal properties of animals figure more prominently in inferences about category membership -- how it is to be classified -- than do conceptions of environment and nurture.⁴¹

If psychological essentialism can be established, then from among the many beliefs associated with certain types of entities, we can experimentally isolate a core group that is most significant for classification. One should note that psychological essentialism does not specify that the only beliefs at the core of the structure of a concept, even for natural kind concepts, are about essences. Psychological essentialism marks only the beliefs especially important for one type of conceptual ability, classification. Similar research might be carried out for other abilities associated with concepts, such as identification and recognition, and it might reveal specific types of core beliefs connected with these abilities as well.

In addition, not every classificatory concept has the sort of psychological essentialist structure we have encountered. Functional concepts, for example, are not likely to involve supposition of an internal essence that is causally responsible for surface properties. But at the core of the structure of such concepts we might well find a belief about a functional definition or about a corresponding placeholder conception. At the core of the concept 'carburetor' we might discover the belief that a carburetor is a device that mixes air with gasoline for internal combustion, and for those who cannot produce such a definition, we might find the belief that nevertheless there is a definition for a carburetor and that mechanics know what it is.

But still, psychological essences of the sort that Gelman and Wellman have investigated

provide a paradigm for individuating the cognitive value of concepts by structural features. By appealing to similarities and differences in the core group of beliefs associated with certain concepts we can judge the cognitive value of various instantiations of these concepts to be the same or different. The kinds of tests Gelman, Wellman and others have developed can be used to confirm hypotheses about similarity in conceptual structure both within persons over time and across different persons. Psychological essentialism, therefore, provides us with a good example of the type of structure a concept may have, a model for individuating the cognitive value of concepts by such structure, and empirical procedures for determining structural intra- and interpersonal similarity.

Further reflection indicates how individuation of the cognitive value of concepts by their structural features might illuminate the three difficulties Bilgrami raises for theories that individuate content by associated beliefs: first, the problem of finding a method for avoiding individuation of content that is too fine-grained, and, more broadly, of finding a method for individuating content in just the right way; second, the specter of losing the distinction between change in belief and change in meaning; and third, the threat of reintroducing analyticity.

On the first issue, individuation by conceptual structure remedies functional role theory's lack of a principled and satisfying method for classifying token-distinct psychological states as having the same content. We might endorse, for example, the principle that beliefs about essential properties carry more weight for individuation than do beliefs about accidental properties, as well as similar principles derived from a study of identification and recognition. Suppose, then, that two scientists are thoroughly knowledgeable about arthritis, but yet the functional roles of their 'arthritis' concepts differ somewhat. We can nevertheless attribute the same concept to them at least in part because they agree on the defining properties of arthritis, and they use a similar set of core beliefs to identify and recognize arthritis.

Individuation by conceptual structure can also help account for the intuition that in Burge's case, psychological explanation could require that Alfred and the expert be attributed

distinct 'arthritis' concepts. Suppose that Alfred has some beliefs about arthritis, but does not know its defining properties. He also has a placeholder conception constituted by the belief that arthritis is a disease characterized by whatever internal properties explain the symptoms of what people in his society call "arthritis". In addition, he has some false beliefs about arthritis, and he lacks a significant measure of the beliefs that would enable him to identify and recognize arthritis at the expert level. In this situation, the difference between the core structures that function in Alfred's and the expert's conceptions of arthritis might well be significant enough to constitute a distinction in their concepts' cognitive value. Individuation by conceptual structure might, therefore, distinguish between concepts' contributions to content where some externalist views would prescind from such a difference. And thus this method potentially provides a better account of certain key intuitions about the kinds of generalizations we need to make in psychological explanation.

In addition, individuation by conceptual structure can also account for the intuition that in Burge's case, psychological explanation could require that Alfred and his counterpart be attributed the same 'arthritis' concepts. Suppose again that Alfred and his counterpart do not differ with respect to any non-intentional individualistic feature. Both believe that the word 'arthritis' refers to certain rheumatoid ailments not only of the joints but also of the muscles and tendons. Both lack a significant proportion of the beliefs that would enable them, in their respective societies and environments, to classify, identify, and recognize arthritis at the expert level. And each has a placeholder conception that he would express as "a disease characterized by whatever internal properties explain the symptoms of what people in my society call "arthritis"." Under these circumstances, individuation by structural features of concepts would authorize attribution of the same 'arthritis' concepts to each, despite the fact that the standards of ordinary attribution practice would prevent us from doing so.

But how could an attribution practice indicate the similarity in their belief contents? It would be misleading to do so by attributing to each the unqualified belief that he has arthritis in

his thigh. For as we have argued, although attributing to Alfred the belief that he has arthritis in his thigh would be warranted by ordinary practice, it might well be mistaken to take it as an accurate report of his attitude. To capture Alfred's concept precisely, we might need to hedge our attribution as we just did, by specifying that although Alfred has such and such a placeholder conception, he lacks a significant measure of the beliefs at the core of the 'arthritis' concept.' And thus the similarity in their attitude-contents would not be captured accurately by attributing to each the unqualified belief that he has arthritis in his thigh.

But perhaps attributing to each a qualified 'arthritis' concept, in a single language, might indicate the similarity accurately enough. For the purposes of communication of evidence about the world it would not be pragmatically valuable to attribute to Alfred or to his counterpart a concept in a language that the audience -- typically his own community -- does not speak. But if we prescind from that purpose, the peculiarity vanishes. For the purpose of psychological explanation, we may need to attribute the same concept to Alfred and his counterpart, and there could well be different languages by which to characterize that concept.

These claims have implications for the truth-conditions of attitude attributions. In Burge's view, the content of Alfred's belief, which one might ordinarily express by 'that he has arthritis in his thigh,' makes a contribution to a truth-condition that is context-dependent only for the indexical components specified by 'he' and 'his.' But if we suppose that the beliefs of Alfred and his counterpart have the same content, and that Alfred's belief is false whereas the counterpart's is true, then the truth-condition of Alfred's belief cannot be context-dependent only to this degree. For nothing else besides a context-dependent element of the truth conditions of the two beliefs could make the difference in their truth-values, and what makes the difference in these truth values is independent of the contribution of the indexicals 'he' and 'his.' It clearly would have to be the contribution to truth conditions of 'arthritis,' or whatever words or phrases occupy the position of this term in an accurate attribution, that makes for the additional context-dependence in the truth conditions.

A good candidate for a context-sensitive component in Alfred's 'arthritis' concept is the placeholder conception. In the placeholder conception 'a disease characterized by whatever internal properties explain the symptoms of what people in my society call "arthritis"' there is indeed a context-sensitive element. For Alfred, the description 'what people in my society call "arthritis"' will refer to the disease called "arthritis" in English, whereas for the counterpart this description will refer to the disease called "arthritis" in twin-English. Or if the placeholder conception makes explicit reference to experts, in Alfred's conception the term 'the experts' will refer to the experts on earth, whereas in the counterpart's conception the phrase 'the experts' will refer to the experts on twin earth. Hence, concepts that in Burge's scheme do not introduce context-dependence into truth conditions do introduce such context-dependence in the view we are considering.

Let us now turn to the second difficulty Bilgrami raises. If a concept is to retain its cognitive value despite change in associated beliefs, some anchor must be available to provide stability for cognitive value when such change occurs. The reason that functional role theories have failed to supply such anchors is that the required distinctions among the beliefs associated with a concept -- distinctions that would allow some of these beliefs to play a more prominent role than others in determining the cognitive value of a concept -- have not been made in a principled way. But psychological essentialism's illumination of classification, and similar results for identification and recognition, would allow for principled distinctions based on conceptual structure. If we individuate the cognitive value of a concept by core beliefs, then it can remain the same despite change in the beliefs other than those at the core. Change of cognitive value would occur only when change occurs in the beliefs at the core of the concept's structure.

The third problem Bilgrami points out is that individuating the cognitive value of a concept by the beliefs of the subject risks making certain concepts analytic of that concept. And indeed, individuation in accord with structural features like those posited by psychological essentialism entails this result, for concepts involved in beliefs found at the core of a concept

would constitute its psychological meaning, and would, in one sense, be analytic of it. But as Hilary Kornblith points out, this admission does not contravene the Quinean opposition to analyticity.⁴² Quine's opposition focusses on the connection between analyticity and apriority, but the type of investigation into the structural core of a concept we have considered is decidedly empirical. In the view we have been developing, conceptual analysis is a significant and fruitful project, but it does not manifest the objectionable features that are challenged by Quine's arguments.

Thus we can see how an appeal to conceptual structure might solve the kinds of problems that arise for a theory of content in which individuation by associated beliefs plays a central role. An account of this kind provides criteria for conceptual identity and difference, allows us to distinguish between change of psychological meaning and change of belief, and shows how a theory of content can accommodate a plausible notion of analyticity. A rejection of recent externalist views about content, therefore, might well be compatible with a psychology that is plausible, principled, and empirical.

Let us now consider two objections to the view we have been developing. First, one might argue that the belief attributions made in the psychological essentialist experiments themselves have as purposes both psychological explanation and communication of evidence about the world. And thus these attributions cannot be assumed to indicate psychological content unequivocally. Much that may seem to be of psychological interest in these attributions may subserve the goal of communication of evidence about the world, and not of psychological explanation, and hence, by my own conjecture, they may deflect the aim at truth about psychological content. Consequently, I have not shown that a theory of psychological content can avail itself of the claims about conceptual structure made by theorists like Gelman and Wellman.

But this objection is mistaken. As became clear in our discussion of Kripke's puzzle, our attributions might aim (mainly) at one goal rather than the other, and differences in purpose will

result in differences in attribution. Since the dominant purpose of the psychological essentialist experiments is to provide instruments for explaining inference and behavior, and not to communicate evidence about the objects used in the experiments to an audience, we have reason to believe that the attributions in the experiments are aimed at specifying psychological content. Moreover, it is difficult to see how the crucial features of the experiments, for example, the attribution to the children of the view that internal properties of animals are more important than surface properties for being what they are, might be understood as a function of the aim to communicate evidence about these animals to an audience. The implausibility of this conjecture provides further reason to take the crucial attributions in the psychological essentialist experiments to illuminate real psychological content.

A further objection derives from an anti-individualist argument for prototype theory developed by Bernard Kobes.⁴³ Kobes points out that, in general, psychology aims to explain not only behavior, but also a subject's classificatory successes and failures. But, he argues, this normative dimension cannot be facilitated if a concept's range of correct application is explained solely in terms of the individual subject's articulations and dispositions. A subject's articulations and dispositions are often insufficient for specifying the range of correct application of the concepts she has, while she is yet committed to there being such a range of correct application. This background commitment, then, should lead us to fix the range of her concept's application non-individualistically: "it is fixed in part by the subject's natural and social environment."⁴⁴

This objection transfers easily to the psychological essentialist view we have been considering: although the articulations and dispositions of the subjects in the psychological essentialist experiments are insufficient to specify the range of correct application of their concepts, these subjects are yet committed to there being such a range. We must therefore individuate the subjects' concepts non-individualistically, and we are thus led back to an externalist account of the sort Burge has advocated. But, in reply, there is an account of the range of correct application of the relevant concepts for which the articulations and dispositions

of the subject do suffice. For the commitment of the subjects to a range of correct application, and the deference to experts and to the world that this commitment involves, will be reflected in such articulations and dispositions. This commitment can be explained by the sort of placeholder conception we have been discussing. A placeholder conception does not specify the range of correct application in the way that a completed science would, but rather with the aid of conceptions like 'whatever the experts believe to cause the symptoms of the disease called "arthritis"' or 'whatever internal properties really cause the symptoms of the disease called "arthritis."' To discover the extension of a placeholder conception we must certainly examine the world beyond the individual, just as is the case for conceptions like 'the first person born in the twentieth century' and 'that man over there.' But we have encountered no reason to think that a placeholder conception itself is to be individuated with reference to factors beyond the individual subject's articulations and dispositions.

Let me make three points of clarification about the account as it has been developed so far. First, individuation of content by structural features may serve as a welcome supplement to functional role theory. Such a view about individuation might well supply the principled criteria for sameness of content that functional role theories require. But one might also accept individuation by structural features and reject functional role theory if one were to deny that content is solely a function of relations to perceptual inputs, behavioral outputs, and other states.⁴⁵ Second, it is consistent with this account, which focusses on the psychological explanation of inference and behavior, that there are other branches of psychology for which attributions of content should be externalist in one or more of the senses we have discussed. Perhaps the psychology of perception is one of those branches, but I shall not take a stand on this issue here.⁴⁶ Third, for all I have argued so far, there may be externalist factors relevant to content attribution even when psychological explanation of inference and behavior is at issue. I have argued that for the psychological explanation of such phenomena, certain prominent externalisms may well be inappropriate, and that in order to specify sameness and difference of content, one might turn

instead to structural features of concepts. This leaves room for the appropriateness of other sorts of externalist characterizations of content for this type of psychology. But I shall not pursue this issue any further here.

VI

In summary, several prominent externalist theories of content conflict with some of our intuitions about psychological explanation. And thus, although individuation of content in accord with these externalisms could suffice for most epistemic and psychological purposes, such views might well be inadequate when certain recalcitrant cases of psychological explanation are at issue. Furthermore, functional role theories seem to lack any principled and satisfying method of classifying token-distinct psychological states as having the same type of content. Recent exploration of the structural features of concepts, however, potentially supplies the kinds of resources needed to individuate content in a way that yields the generalizations in psychological explanation we want to make. At this stage a theory of content individuation along these lines can be neither fully spelled out nor confirmed. But such a theory does underwrite an ongoing empirical research program that will ultimately determine its prospects.⁴⁷

NOTES

1. Hilary Putnam, "The meaning of 'meaning'," in Hilary Putnam, Philosophical papers, v. 2 (Cambridge: Harvard University Press, 1975), pp. 215-271; Saul A. Kripke, Naming and Necessity (Cambridge: Harvard University Press, 1980); "A Puzzle About Belief," in Meaning and Use, ed. A. Margalit (Dordrecht, D. Reidel, 1979); Tyler Burge, "Individualism and the Mental," in Midwest Studies in Philosophy v. VI, eds. P. French, T. Uehling, H. Wettstein (Minneapolis: University of Minnesota Press, 1978), pp. 73-121; see also, his "Other Bodies," in Thought and Content, ed. A. Woodfield (Oxford: Oxford University Press, 1979), pp. 97-120; "Two Thought Experiments Reviewed," Notre Dame Journal of Formal Logic, 23 (1982), pp. 284-293; "Individualism and Psychology," Philosophical Review 95 (1986), pp. 3-45; "Individuation and Causation in Psychology," Pacific Philosophical Quarterly 70 (1989), pp. 303-322; Lynne R. Baker, Saving Belief (Princeton: Princeton University Press, 1987); Ruth Millikan, Language, Thought, and Other Biological Categories (Cambridge: MIT Press, 1984), "Thoughts Without Laws: Cognitive Science Without Content," Philosophical Review 95 (1986), pp. 47-80; Daniel Dennett, "Evolution, Error, and Intentionality," in The Intentional Stance (Cambridge: MIT Press, 1987), pp. 287-321.

2. Ned Block, "An Advertisement for a Semantics for Psychology," in Midwest Studies in Philosophy, v. X, ed. P. French, T. Uehling, H. Wettstein (Minneapolis: University of Minnesota Press, 1986), pp. 615-678; Brian Loar, "Conceptual Role and Truth Conditions," Notre Dame Journal of Formal Logic 23 (1982), pp. 272-283, "Social Content and Psychological Content," in Contents of Thoughts, ed. R. H. Grimm and D. D. Merrill (Tucson: The University of Arizona Press, 1985), pp. 99-110. In some functional role theories content does not supervene on the physical features of the individual; see, for example, Gilbert Harman, "Functional Role Semantics," Notre Dame Journal of Formal Logic 23 (1982), pp. 242-256, and "Wide

Functionalism," in The Representation of Knowledge and Belief, eds. R. Harnish and M. Brand (Tucson: University of Arizona Press, 1986).

3. Ned Block, "An Advertisement for a Semantics for Psychology;" Jerry Fodor, "Banish DisContent," in Language, Mind, and Logic, ed. J. Butterfield (Cambridge: Cambridge University Press, 1986), reprinted in Mind and Cognition, ed. William G. Lycan (Oxford: Blackwell, 1990), pp. 420-438, "Meaning Holism," in Psychosemantics (Cambridge MIT Press, 1987), "Substitution Arguments and the Individuation of Beliefs," in A Theory of Content and Other Essays (Cambridge: MIT Press, 1990), pp. 161-176, at pp. 161-2.

4. See, for example, Brian Loar, "Social Content and Psychological Content;" Akeel Bilgrami, Belief and Meaning (Oxford: Blackwell, 1992), pp. 1-63.

5. David Kaplan, "Demonstratives," in Propositions and Attitudes, eds. Nathan Salmon and Scott Soames (Oxford: Oxford University Press, 1979).

6. Jerry Fodor, "Substitution Arguments and the Individuation of Beliefs," 167-8.

7. See, for example, Fodor, "Banish DisContent" and Bilgrami, Belief and Meaning, pp. 30-1.

8. Gottlob Frege, "On Sense and Reference," in Translations of the Philosophical Writings of Gottlob Frege, eds. P. Geach and M. Black (Oxford: Blackwell, 1966).

9. Saul A. Kripke, Naming and Necessity; "A Puzzle About Belief," at pp. 241-2.

10. Ibid. pp. 241-2.

11. Brian Loar, "Social Content and Psychological Content."

12. Bilgrami, Belief and Meaning, 16-19.

13. Tyler Burge, "Sinning Against Frege," Philosophical Review (1979), pp. 398-432.

14. Tyler Burge, "Individualism and the Mental."

15. In "Social Content and Psychological Content" Loar argues that a Fregean puzzle reinforces these sorts of problems for Burge's anti-individualist theory of cognitive value. The following case illustrates Loar's contention. Suppose that Alfred is French, and he believes 'J'ai de l'arthrite à ma cuisse,' and we therefore attribute to him the belief that he has arthritis in his thigh, using our concept of arthritis -- a concept which we also attribute to the experts. On one occasion he travels to England, where he learns, in English, the meaning of the experts' conception of 'arthritis,' including that it refers to a disease of the joints. Reflecting on the pain in his thigh, and the fact that the thigh is not a joint, he comes to believe the English sentence 'I do not have arthritis in my thigh.' But he maintains the belief 'J'ai de l'arthrite à ma cuisse' for many months, not realizing that arthrite = arthritis. But then one day he makes the connection that arthrite = arthritis, which is informative for him. This belief has profound effects on his inferences, the first of which is the abandonment of the belief 'J'ai de l'arthrite à ma cuisse.' This is intended to show that the cognitive value of the expert's concept is distinct from that of Alfred's original 'arthritis' concept. One might imagine a similar example, all in English, in which for a time after Alfred learns from the experts that arthritis is a disease of the joints, he believes that there are two diseases called arthritis. (Perhaps he believes that 'arthritis' is like 'fruit,' a word with two distinct definitions whose extensions overlap. By both definitions apples are fruits, but peas are so by only one.) But then he comes to believe that arthritis = arthrite, which for him, at that time, is

informative.

This type of Fregean puzzle does not, however, help to show that the cognitive value of Alfred's and the expert's concepts are distinct. One should note that it departs from the form of Frege's original puzzle, since it aims to demonstrate an interpersonal rather than an intrapersonal difference in cognitive value -- a difference in cognitive value between Alfred and the expert rather than within Alfred. But such extension of Frege's test would draw distinctions in cognitive value where they do not exist. Consider this counterexample. Suppose Alfred's French counterpart has precisely the same beliefs about arthritis that Alfred has, except that the counterpart would express them in French, and that the inference and behavior associated with these 'arthritis' concepts are as similar as they can be for people who speak different languages. We would want to make psychological generalizations that regard these concepts as having the same cognitive value. But we could imagine Alfred travelling to France, learning his counterpart's 'arthrite' beliefs, believing, for a time, that he has two different ailments in his thigh, and later coming to the informative realization that arthritis = arthrite. Hence, this interpersonal Fregean test would illegitimately distinguish the concepts of Alfred and his counterpart.

16. Tyler Burge, "Sinning against Frege," at p. 419; see also his "Individualism and Psychology," p. 7, and "Belief De Re," Journal of Philosophy 56 (1978), pp. 338-362.

17. John Perry, "Belief and Acceptance," Midwest Studies in Philosophy V (Minneapolis: University of Minnesota Press, 1980), pp. 533-542; Donald Davidson, "Knowing One's Own Mind," Presidential Address, Pacific Division Meeting of the American Philosophical Association, March 28, 1986, Proceedings of the American Philosophical Association, pp. 441-458, at p. 449; Michael Devitt, "A Narrow Representational Theory of the Mind" in Mind and Cognition, ed. William G. Lycan (Oxford: Blackwell, 1990), pp. 371-398.

18. Devitt, p. 389.

19. "A Puzzle About Belief," p. 259.

20. In the "unarticulated constituent" account of Mark Crimmons and John Perry ("The Prince and the Phone Booth: Reporting Puzzling Beliefs, The Journal of Philosophy 86 (1989), pp. 685-711) and in the "hidden indexical" theory of Stephen Schiffer ("Belief Ascription," The Journal of Philosophy 89 (1992), pp. 499-521), ordinary belief attributions are construed as always making implicit reference to modes of presentation. Thus 'Pierre believes that London is pretty' specifies not simply a two-place relation between Pierre and the proposition that London is pretty, but a three-place relation among Pierre, the proposition, and an unarticulated mode of presentation under which Pierre believes the proposition. If the unarticulated modes of presentation for Pierre's two beliefs are different, then we have a solution to Kripke's puzzle. These semantical accounts are compatible with my views, but a thorough examination of essentially semantical issues is beyond the scope of this article.

21. See, for example, "Individualism and Psychology," p. 7.

22. Fred Dretske, Knowledge and the Flow of Information (Cambridge: MIT Press, 1981); Ruth Millikan, Language, Thought, and Other Biological Categories (Cambridge: MIT Press, 1984), "Thoughts Without Laws: Cognitive Science Without Content," Philosophical Review 95 (1986), pp. 47-80; Jerry Fodor, A Theory of Content and Other Essays (Cambridge, MIT Press, 1990), pp. 51-136; see also Mohan Matthen, "Biological Functions and Perceptual Content," Journal of Philosophy 85 (1988), pp. 5-27.

23. Jerry Fodor, A Theory of Content and Other Essays, pp. 59-61.

24. Ruth Millikan, "Thoughts Without Laws, Cognitive Science Without Content," pp. 67-71.

25. Fodor provides a splendid catalogue of these problems in his A Theory of Content and Other Essays, pp. 51-136. See also William Seager, "Fodor's Theory of Content," Philosophy of Science 60 (1993), pp. 262-277.

26. Bilgrami, Belief and Meaning, pp. 58-60.

27. Jerry Fodor, "Substitution Arguments and the Individuation of Beliefs."

28. Ibid, pp. 168, 174-5.

29. Ruth Millikan, "Thoughts Without Laws, Cognitive Science Without Content," pp. 63-67.

The view cited here is similar to Fodor's refinement of Millikan's idea in A Theory of Content (pp. 69-70). Daniel Dennett remarks, citing Millikan "... so evolution can select an organ for its capacity to oxygenate blood, can establish it as a lung. And it is only relative to such design "choices" or evolution-"endorsed" purposes -- raisons d'être -- that we can identify behaviors, actions, perceptions, beliefs, or any of the other categories of folk psychology" ("Evolution, Error, and Intentionality," in The Intentional Stance (Cambridge: MIT Press, 1987), pp. 287-321, at p. 300).

30. Jerry Fodor, A Theory of Content and Other Essays, p. 77.

31. Richard Boyd, "Materialism Without Reductionism: What Physicalism Does Not Entail," in Readings in the Philosophy of Psychology, v. 1, ed. Ned Block (Cambridge: Harvard University Press, 1980), pp. 67-106; Jerry Fodor, "Special Sciences," in Readings in the Philosophy of

Psychology, v. 1, pp. 120-133; Hilary Putnam, "The Nature of Mental States," in Readings in the Philosophy of Psychology, v. 1, pp. 223-231, "Philosophy and Our Mental Life" in Readings in the Philosophy of Psychology, v. 1, pp. 134-143; Derk Pereboom and Hilary Kornblith, "The Metaphysics of Irreducibility," Philosophical Studies 63 (1991), pp. 125-145.

32. Stephen Stich, "What is a Theory of Mental Representation?" Mind 101 (1992), pp. 243-261; Michael Tye, "Naturalism and the Mental," Mind 101 (1992), pp. 421-441.

33. Bilgrami, Belief and Meaning, pp. 10-13.

34. Ibid, pp. 10-13, 83-129.

35. In one Kantian sense, a judgment is analytic in virtue of concept-containment; cf. Critique of Pure Reason, A6/B10-A10/B14.

36. Bilgrami, Belief and Meaning, pp. 122-9.

37. Ibid, pp. 83-113.

38. Douglas Medin and Andrew Ortony, "Psychological Essentialism," in Similarity and analogical reasoning, eds. S. Vosniadou and A. Ortony (Cambridge: Cambridge University Press, 1989) pp. 179-195; Douglas Medin, "Concepts and Conceptual Structure," American Psychologist, v. 44, n. 12, pp. 1469-1481.

39. For a notion similar to the placeholder conception, see Georges Rey's "Concepts and Stereotypes: A Reply to Smith, Medin, and Rips," Cognition 19 (1985), pp. 197-303.

40. Susan A. Gelman and Henry M. Wellman, "Insides and essences: Early understandings of the non-obvious," Cognition 38 (1991), pp. 213-244, at pp. 223-229.

41. Ibid, pp. 229-234.

42. Hilary Kornblith, Inductive Inference and Its Natural Ground (Cambridge: MIT Press, 1993), p. 80.

43. Bernard W. Kobes, "Semantics and Psychological Prototypes," Pacific Philosophical Quarterly 70 (1989), pp. 1-18.

44. Ibid, p. 10. Dispositions and articulations of the sort that Kobes discusses, and an anti-individualist argument based on them, are examined in depth in Tyler Burge's "Wherein is Language Social," in Propositional Attitudes: the Role of Content in Logic, Language, and Mind, eds. C. Anthony Anderson and Joseph Owens (Stanford: Center for the Study of Language and Information, 1990), pp. 113-130, esp. 122-130.

45. Derk Pereboom, "Why a Scientific Realist Cannot Be a Functionalist," Synthese 88, 1991, pp. 341-358.

46. See Tyler Burge, "Individualism and Psychology;" Bernard W. Kobes, "Individualism and Artificial Intelligence" in Philosophical Perspectives 4: Action Theory and Philosophy of Mind, ed. James Tomberlin (Atascadero, CA: Ridgeview Publishing Company, 1990); "Cartesian Error and the Objectivity of Perception," in Contents of Thoughts, ed. R. H. Grimm and D. D. Merrill (Tucson: The University of Arizona Press, 1985); Gabriel Segal, "Seeing What is Not There," The Philosophical Review 98 (1989), pp. 189-214; Frances Egan, "Must Psychology Be

Individualistic?" The Philosophical Review 100 (1991), pp. 179-203.

47. I am very grateful to Lynne Rudder Baker, David Christensen, Hilary Kornblith, and Bernard Kobes for enlightening commentary and discussion.