

Dispositions Indisposed: Semantic Atomism and Fodor's Theory of Content¹

(Appears in *Pacific Philosophical Quarterly*, September, 2000)

Robert D. Rupert

Abstract: According to Jerry Fodor's atomistic theory of content, subjects' dispositions to token mentalese terms in counterfactual circumstances fix the contents of those terms. I argue that the pattern of counterfactual tokenings alone does not satisfactorily fix content; if Fodor's appeal to patterns of counterfactual tokenings has any chance of assigning correct extensions, Fodor must take into account the contents of subjects' various mental states at the times of those tokenings. However, to do so, Fodor must abandon his semantic atomism. And while Fodor has recently qualified his atomism, the cognitively holistic nature of dispositions continues to undermine his view.

I. Introduction

Beliefs seem to mediate the application of humans' mental representations of natural kinds.² For example, whether a physicist is likely to think "there goes a proton" at a given time depends greatly on beliefs about physics acquired during her scientific training. Jerry Fodor recognizes that mental states causally influence our tokenings of natural kind terms in mentalese but nonetheless denies that the *content* of these states plays an ineliminable role in fixing the reference of natural kind terms in mentalese:³

But though protons typically exert causal control over 'protons' via the activation of intentional mechanisms, a naturalistic semantics doesn't need to specify all

that. All it needs is that the causal control should actually obtain, *however* it is mediated.⁴

More recently Fodor has made similar remarks about what he calls the “engineering” task of nomically linking a mentalese term with its extension: “Answers to this engineering question can unquestion-beggingly appeal to the operation of semantic and intentional mechanisms, since ‘semantic’ and ‘intentional’ are presumed to be independently defined.”⁵ According to Fodor, so long as we characterize the nomic relation that grounds reference without invoking content or other intentional notions, it is irrelevant that content-laden states often play a significant role in sustaining the reference-fixing nomic relation. This view of the role of mental states in the grounding of reference seems to cohere well with a further position of great importance to Fodor: semantic atomism, i.e., the view that the content of each mentalese term is fixed independently of the content of all other terms.⁶ In this paper, I argue that, appearances to the contrary, Fodor cannot plausibly embrace both semantic atomism and his theory of content for natural kind terms in mentalese.

In explicating his theory of content, Fodor often invokes, as determinants of reference, subjects’ dispositions to token mentalese terms in counterfactual circumstances;⁷ and while he sometimes expresses discomfort at having to make such appeals,⁸ he has got himself into a spot. Until Fodor offers some other way to understand the nature of the nomic connections involved in reference-fixing, his atomism remains untenable: dispositions will determine acceptable extension assignments only if they operate in tandem with the contents of the mental states that mediate the tokening of natural kind terms in mentalese.

If, as my argument suggests, the *contents* of subjects’ intentional states play an essential role in the fixation of reference for natural kind terms in mentalese, we may suspect that naturalistic

semantics for natural kind terms is hopeless: naturalistic semantic theory is saddled with holism and thus unable to discharge all intentional idioms in explaining how the content of a given natural kind term is fixed.⁹ I do not draw the antinaturalistic conclusion, for my argument leaves open naturalistic options that do not invoke Fodorean dispositions as determinants of content; I return to this issue briefly in closing.

II. Fodor's Dispositions

When canvassing objections to his asymmetric dependence theory of content,¹⁰ Fodor's imaginary conversant challenges him to "*Do the Twin Cases.*"¹¹ The twin cases to which Fodor's interlocutor refers are those originally found in Hilary Putnam's writings on the reference of natural kind terms in natural languages. Imagine that far away, in a distant galaxy, there exists a planet exactly like Earth; call it 'Twin Earth'. In respect to local facts, the sole difference between Earth and Twin Earth is that whereas terrestrial lakes and rivers are filled with H₂O (plus impurities), no such substance exists on Twin Earth; wherever H₂O exists on Earth, Twin Earth instead contains, in the corresponding geographical position, a colorless, tasteless, thirst-quenching liquid made up of mystery compound XYZ. This mystery compound is so much like H₂O that the average dwellers of Earth and Twin Earth would not be able to tell the two substances apart were they confronted with samples of both. Any English speaker will do here, but take a particular, Earth-dwelling English speaker, Oscar; on Twin Earth he has a counterpart, call him 'twin-Oscar', who speaks twin-English. Oscar and twin-Oscar are exactly alike in all physical respects: look at them however closely you like inside or out, you will find no difference (except for the differential presence of H₂O and XYZ in Oscar's and twin-Oscar's organic profiles, but I ignore this complication). When Oscar says 'water is wet', twin-Oscar utters the acoustically identical sentence 'water is wet'. However, despite their two sentences'

possessing identical acoustical properties, and despite twin-Oscar's being a molecule for molecule duplicate of Oscar, each speaks of a different substance when he says 'water is wet': Oscar speaks of H₂O, twin-Oscar of XYZ.¹²

Putnam famously drew on Twin Earth examples to support his causal theory of reference for natural kind terms in a natural language. Yet, although philosophers routinely extend the moral of Putnam's story, viz. that "meaning ain't in the head," to the case of thought content, the two cases differ importantly. In what follows, then, I begin with Fodor's explanation of how reference is fixed for natural kind terms in natural language, proceeding directly to the case of natural kind terms in mentalese:

The thing to keep your eye on is this: It's built into the way that one tells the Twin Earth story that it's about kind-terms (*mutatis mutandis*, kind-concepts). In particular, it's part of the story about "water" being a kind-term that English speakers intended it to apply to all and only stuff of the same (natural) kind as paradigmatic local samples...My point is that the intention to use "water" only of stuff of the same kind as the local samples has the effect of making its application to XYZ asymmetrically dependent on its applications to H₂O *ceteris paribus*.¹³

Speakers' intentions explain the privileged status of the causal connection between H₂O and 'water'. Such intentions determine in the speaker a disposition to apply 'water' to some substances but not to others; they establish a pattern of utterances across counterfactual situations, and this ultimately fixes the reference of 'water':

Given that people are disposed to treat "water" as a kind-term...they would apply it to XYZ only when they *mistake* XYZ for H₂O; only when (and only because)

they can't tell XYZ and H₂O apart. Whereas, given a world in which they *can* tell XYZ and H₂O apart (and in which their intentions with respect to "water" are the same as they are in *this* world) they will continue to apply "water" to H₂O and refrain from applying it to XYZ.¹⁴

If Oscar could tell H₂O and XYZ apart, he would apply 'water' only to H₂O.¹⁵

Fodor's suggestion that speakers' *intentions* fix the reference of 'water' may cause naturalists discomfort; however, recall the description of Fodor's position with which I began the paper: an intention might in some way be responsible for the speaker's disposition, but it is the disposition itself, characterized nonsemantically, that secures the content-fixing asymmetric dependence. Thus, Fodor handles the twin cases without appealing to the contents of speakers' intentions; he need only appeal to Oscar's and twin-Oscar's verbal behaviors in counterfactual situations.

When Fodor turns explicitly to natural kind terms in mentalese, he is careful to avoid even the appearance of antiatomistic impropriety. Fodor's solution to the twins-puzzles as they arise for natural kind concepts is just the same as his solution in the case of natural language terms, "but for the talk about intentions and policies."¹⁶ Fodor goes on to identify a natural kind concept as a mentalese term that "tracks" a natural kind; to say that a subject's mentalese term tracks a kind is to attribute to the subject a certain reference-fixing disposition: in counterfactual circumstances where the subject can distinguish between the relevant kind members, she tokens the mentalese term in question in response to the kind members that make up its extension, but not in response to members of other natural kinds.

Students of Fodor's work may wonder, at this point, whether I have simply ignored what seem to be Fodor's pertinent protestations. Fodor admits to being uncomfortable with the talk

about counterfactuals, claiming to use such talk only for want of some better way to talk about nomic connections:

I suspect, in particular, that some of the troubles we're about to survey stem not from there being anything wrong with the proposal that content rests on asymmetrical dependences among nomological relations, but rather from there being everything wrong with the assumption that claims about nomological relations need counterfactual/possible world translations.¹⁷

Fodor's reservations are duly noted. But how, then, should we understand what it is to be nomically connected, if not in terms of counterfactual supporting generalizations--in the present context, generalizations about mentalese term tokenings and their causes? Time and again, these are precisely the terms in which Fodor casts his discussion of content-fixing nomic connections. For example, "Where nomic connections are the issue...what counts is only the counterfactuals."¹⁸ And in explaining his diagnosis of Donald Davidson's Swampman thought experiment, Fodor says that Swampman refers to H₂O with his tokenings of 'water' because "it's water that *would* cause his 'water' tokens in the worlds that are closest to the one that Swampman actually lives in."¹⁹ Fodor attributes content to concepts with empty extension also by appeal to counterfactuals: his theory assigns the intentional content *unicorns* to the concept 'unicorn' because "people would apply 'unicorn' to unicorns if there were any."²⁰ Furthermore, bear in mind Fodor's remarks about behaviorism and information-based semantics. Although Fodor roundly rejects behaviorism in general, in *TOC* he praises a Skinner-inspired semantics (of a sort of which, Fodor notes, Skinner himself would most likely have disapproved) for yielding just the sort of semantic theory Fodor wants--one that allows intentional states to influence term-

tokening, within the theoretical framework of atomism;²¹ Fodor equates the new Skinnerian approach, counterfactuals and all, to the idea behind Fodor's own view, the idea that content is grounded in nomic connections.²² While Fodor may, on some level, be aware of the attendant risks, he seems to see no viable option to explaining content-fixing nomic connections in terms of the subject's disposition to token mentalese terms in counterfactual situations. This comes as no great surprise, given that (1) Fodor claims content to be grounded in nomic connections and (2) being a counterfactual-supporting generalization appears to be the most promising idea on offer of what it is to be a nomic connection.

Here we arrive at a second point of record that may cast doubt on my disposition-based exegesis of Fodor's theory of content. In *Concepts*, Fodor seems to distance his view from any that relies heavily on claims about subjects' abilities to discriminate between members of different kinds. He says, "I am...not committed to construing locking in terms of a capacity for discriminated responding (or, indeed, of anything epistemological). Locking reduces to nomic connectedness," he says.²³ If Fodor is deeply committed to a disposition-based explication of his theory, as I have argued, why does he openly spurn "capacit[ies] for discriminated responding"? Here I take Fodor to be addressing matters that do not bear directly on the nature of the *counterfactual* responses constitutive of content-fixing dispositions. Fodor simply does not want to give the impression that being locked to a property implies any particular range of discriminatory skills *in the real world*.²⁴ Thus, Fodor's point about discriminatory capacities in *Concepts* does not conflict with a disposition-based interpretation of Fodor's theory of content: if your mentalese term 'water' tracks H₂O, then in the nearest worlds where you do happen to have certain discriminatory capacities, i.e., you can discriminate between H₂O and XYZ, you will

apply ‘water’ to H₂O and not to XYZ; but this does not imply that *at present* you possess any particular discriminatory capacities.

Fodor would like to solve Twin Earth puzzles without having to accord any reference-fixing role to the contents of subjects’ intentions. Fair enough, but Fodor seems stuck with dispositions doing the work of subjects’ intentions to treat natural kind terms in mentalese as such. Below I argue that Fodor’s atomism is inconsistent with his appeal to dispositions, the difficulty arising from the cognitively holistic nature of dispositions. Before proceeding to the critical discussion, however, I address one last worry that may arise in the minds of careful readers of *TOC*. Thus far, I have considered only what Fodor calls in *TOC* the “purely informational” version of his theory of content, which solves Twin Earth puzzles by appeal to nomic-connections-cum-dispositions alone. But in *TOC*, Fodor also presents, and at some points seems to favor, what he calls a “mixed” view, one that appeals, not only to lawlike connections as determinants of content, but also to the actual history of the subject in question.²⁵ To render his theory mixed, Fodor incorporates a restriction that says, in effect, that if a term is to refer to the members of a given kind, members of that kind must have actually caused the tokening of the relevant mentalese term in the subject in question. On such a view, it is clear why Oscar’s concept ‘water’ does not refer to XYZ, regardless of what dispositions he might have, for XYZ has never caused the tokening of ‘water’ in Oscar or in any other Earthling. Why, then, would I take the time to stress the failings of the purely informational view, when a switch to the mixed theory absolves Fodor’s semantic theory of these faults?

Here emerges the broader theme of the present work. In *TOC*, Fodor found himself struggling with a question of central theoretical importance in naturalistic semantics: Should content be assigned on the basis of the subject’s (or her species’) actual history of tokening, or

rather on the basis of the subject's disposition to token mentalese terms in counterfactual circumstances? I contend that in resolving his internal struggle, Fodor chose the wrong side--generally speaking, not simply the wrong version of his asymmetric dependence theory--the purely dispositional side; this is the position he now holds. Thus, I intend the following sections to constitute a principled critique of any naturalistic theory that assigns content to natural kind terms on the basis of subjects' nonsemantically individuated dispositions to token under counterfactual circumstances.

III. Intentional States and the Individuation of Dispositions

It is a serious problem for behaviorist accounts of language learning that linguistic behavior is controlled by a complex nexus of factors, rather than by simple associations of responses with stimuli. In particular, critics have chided behaviorist theories for their inability to account for the effects of a subject's mental states on her disposition to verbally respond. Consider a remark of Fodor in this regard, "Pay me enough and I will refer here and now to anything you like...And enough wouldn't come to much."²⁶ Fodor's quip illustrates how a person's current beliefs, desires, and other mental states affect what she is inclined to say in a given situation. If a person has no desire for money or eschews puerile silliness, then she may not be at all inclined to say "funky turtle phlegm" in response to someone else's offer to pay for that utterance. In contrast, neither excessive pride nor a sufficient lack of interest in money would, I presume, keep Fodor from taking the payoff. (Perhaps for Fodor this has changed some with age; that being the case would only make more pointed the way in which the subject's verbal response depends on which mental states she happens to be in when the stimulus occurs.) The complaint against behaviorism seems decisive, because (1) behaviorism explicitly denies the utility of talk about mental states when giving psychological explanations of linguistic behavior and (2) behaviorists

fail in their attempt to reduce to stimulus-response generalizations those mental states that influence subjects' verbal behaviors.

Fodor is no fan of behaviorism as a theory of linguistic behavior. Fodor's theory of content allows the operation of intentional mechanisms to effect asymmetric dependencies, and thus is clearly not a behaviorist theory. Yet, considerations parallel to those raised against behaviorism seem to apply to Fodor's characterization of subjects' content-grounding dispositions: the way in which Oscar will respond to H₂O and XYZ in situations where he can tell the two apart is a function of the content of various of Oscar's mental states, not the simple result of his being differentially stimulated by H₂O and XYZ. We should doubt, therefore, that the content of 'water' can be fixed in an atomistic fashion by a brute disposition: as behaviorists fail in their attempt to reduce to stimulus-response generalizations the mental states that affect linguistic responses, Fodor cannot ignore Oscar's intentional states in favor of Oscar's dispositions to token.

In what follows I express my concern in two ways; which way of speaking is more apt depends on how Fodor himself chooses to characterize his approach to the semantics of natural kind terms in mentalese. Fodor may wish to describe his view in quasi-Putnamian terms: Oscar has a disposition to treat his concept 'water' *as a natural kind term*, which, qua that sort of disposition, has a special kind of content-fixing effect. The reference-fixing effect of Oscar's disposition, viz. that Fodor's theory of content assigns a natural kind (H₂O) as the extension of 'water', determines Oscar's disposition to be of the special sort had by someone who treats 'water' as a natural kind term.

If Fodor frames his view in the foregoing fashion, my criticism takes the following form: Typical subjects, who seem to be disposed to treat certain of their mentalese terms as natural

kind terms, do not instantiate the appropriate content-fixing dispositions: in the relevant counterfactual circumstances (i.e., relevant from the standpoint of Fodor's theory of content), many otherwise typical subjects do not display the patterns of tokenings that, according to Fodor's theory of content, would fix the reference of the mentalese terms in question to the appropriate natural kinds. In fact, given the variety of responses of a single subject across the relevant counterfactual cases, it would seem that the terms in question entirely fail to track natural kinds, even for individual subjects. We might think it sensible to say that the typical subject's intention to treat a natural kind concept as such appropriately fixes the reference of that term; however, for Fodor to take this insight seriously and assign the right extensions, he must abandon atomism: Fodor seems forced to say that the typical subject has a reference-fixing disposition to treat a given mentalese term as a natural kind term because she displays a pattern of counterfactual tokenings that is appropriate to the *contents* of her current mental states, whatever they happen to be; for Fodor's approach to yield the right results, he must amend his theory so as to circumscribe the counterfactual situations required to stimulate a 'water'-tokening in the typical subject in a way that is sensitive to the *contents* of the subject's mental states.²⁷

Alternatively, Fodor may wish to drop this talk of subjects' dispositions to treat certain mentalese terms *as natural kind terms*. He may simply direct us to apply his theory of content to all mentalese terms and see which extension assignments result.²⁸ Once we have canvassed all of the asymmetric dependencies, if a natural kind stands as the extension of a mentalese term, then we are free to call that term a "natural kind term," but we waste breath talking as if the reference of natural kind terms is fixed in some unique way.

If Fodor chooses this second way of expressing his view, I state my objection differently: If we assign extensions in accordance with the asymmetric dependencies alone, we assign the

wrong extensions to many natural kind terms; furthermore, there seem to be terms that we would take to be natural kind terms to which we must deny such status. In the latter case, I have in mind terms that seem to be natural kind terms but to which the asymmetric dependence theory does not assign stable natural kinds. For some mentalese terms that we take to be natural kind terms, the following would seem to hold: there does not exist a nomic connection between the term in question and a single natural kind upon which depend all other nomic connections that cause the tokening of that term. When we apply the asymmetric dependence theory to some natural kind terms, the result is a hodge-podge of nomic connection-breakings and resulting patterns of term-tokenings. For Fodor to assign the correct extensions, he must allow the *contents* of the subject's various mental states to clear up the mess, determining the correct extension assignments. And to do so, Fodor must abandon his semantic atomism.

Return now to Oscar's mentalese term 'water'. We expect any satisfactory theory of content to assign H₂O as the reference of 'water'. For Fodor, this amounts to saying that Oscar instantiates a disposition: Oscar applies 'water' to H₂O, but not to XYZ, in situations where he can tell the difference between H₂O and XYZ. However, as likely as not, Oscar fails to instantiate this disposition. There are numerous, perhaps an infinite number of, ways in which Oscar may become sensitive to the differential presence of H₂O and XYZ.²⁹ Whether, upon a particular discrimination-facilitating change, Oscar will token 'water' when faced with H₂O depends on intentional facts about Oscar; it depends on his state of mind.

Imagine that XYZ has somehow been brought to earth. Before XYZ is released into the earthly environment, chemists lace XYZ with red dye and a chemical agent that keeps XYZ from mixing with H₂O. To help the folk keep their substances straight (perhaps so that those who wish to can avoid ingesting XYZ), chemists hand out color-coded key cards. In the world as I

have described it, Fodor's theory of content entails that Oscar will apply 'water' to H₂O but not to XYZ; 'water' should track H₂O for Oscar. Yet, whether or not Oscar applies 'water' only to H₂O (the clear liquid) in this case depends on a host of Oscar's other mental states, for example, whether Oscar trusts chemists or instead thinks they are part of that evil conspiracy known as "science." Let us say that Oscar is a member of the American militia movement and thinks that chemists are government agents out to control the minds of the people by putting drugs in the water supply (remember fluoride!); suspicious as he is, Oscar does not, in the circumstances described, token 'water' in response to H₂O. He may well exhibit *exactly the contrary* pattern of tokenings of 'water' to what one would expect on Fodor's simple dispositional view: convinced that the government is out to trick him, Oscar tokens 'water' in response to the red liquid (XYZ) but not in response to the clear liquid (H₂O). This is the case even though Oscar seems to treat, and would seem to have always treated, 'water' as a kind concept. As time goes by, some may wonder whether the reference of Oscar's mentalese term 'water' changes: eventually it may come to refer to XYZ. But surely, when XYZ is first introduced, Oscar's 'water'-thoughts refer to good old H₂O.

I have described a complicated mechanism by which Oscar becomes differentially sensitive to H₂O and XYZ. It is important to note, however, that the point of my example carries over to cases where we imagine only slight changes that render XYZ and H₂O distinguishable to Oscar. No matter what change in the universe we imagine, Oscar may currently possess beliefs that will affect the pattern of his tokenings once the change has taken place; i.e., he may have some beliefs that result in a pattern of tokenings not consistent with the assignment of H₂O to 'water', even though he has, and has had for years, the intention to apply 'water' to all stuff of the same kind as local samples.³⁰

In Fodor's defense, one might think that however difficult it is to *characterize* Oscar's overall disposition to treat 'water' as a natural kind term, the relevant disposition lies present in any complete nonintentional accounting of Oscar. Oscar's physical constitution completely determines his responses in counterfactual circumstances, and his disposition to treat 'water' as a natural kind term is no more than this pattern of such responses; in this way, Oscar's disposition is independent of the content of any of his mental states. This defense of Fodor fails. We are confident that Oscar, as a standard subject, is disposed to treat 'water' as a natural kind concept; however, Oscar's tokenings across counterfactual circumstances, determined as they may be by Oscar's physical constitution, do not determine that he has such a disposition. Make H₂O and XYZ distinguishable to Oscar in one way, and 'water' seems to track H₂O; make the two distinguishable in another way, and Oscar tokens 'water' in response to XYZ but not to H₂O. The pattern of responses itself seems to imply that Oscar does *not* treat 'water' as a natural kind concept. It is only the *contents* of the mental states responsible for Oscar's different responses in different counterfactual circumstances that determine those responses to manifest a disposition to treat 'water' as a kind concept. It is only because of the contents of Oscar's mental states that his pattern of tokenings, determined as it is by his physical constitution, counts as a disposition to treat 'water' as a natural kind term.

Here is my point put solely in terms of Fodor's asymmetric dependence theory. Given initially Oscar's responses to H₂O (he does not token 'water') and XYZ (he tokens 'water'), the expected content-fixing asymmetric dependency does not hold; we have broken the connection between H₂O and 'water' (in Oscar), but Oscar continues to token 'water' in response to XYZ. Nonetheless, the conviction remains that Oscar's mentalese term 'water' refers to H₂O (at least during the time immediately following the introduction of XYZ into Oscar's environment).

Furthermore, when we consider the various ways in which the connection between H₂O and ‘water’ might be broken, it seems that Fodor’s theory of content will not assign a single natural kind to Oscar’s mentalese term ‘water’. Depending on the nomic alteration made, combined with what beliefs and desires Oscar might have, the XYZ-causes-‘water’ connection will sometimes appear to be asymmetrically dependent on the H₂O-causes-‘water’ connection, sometimes not. Without taking into consideration the contents of Oscar’s various mental states that influence his patterns of tokenings, Fodor’s theory of content seems to imply ‘water’ is not even a natural kind concept.

Fodor might retort that the changes in our world I have imagined do not alter the *nomic* connection between H₂O and ‘water’: the nomic connection remains, Fodor might claim, even though Oscar’s mediating beliefs keep him from tokening ‘water’ in response to H₂O. If Fodor chooses this route, blaming Oscar’s mediating beliefs for his failure to exhibit the expected pattern of counterfactual tokenings, then Fodor seems to face a version of Paul Boghossian’s complaint (to be discussed below): Fodor must somehow differentiate, across counterfactual situations, between a subject’s tokenings that are due to nomic connections only (which are to count toward the determination of reference) and those tokenings that result from mediating beliefs (which tokenings are to be discounted); he offers no way of doing so.

IV. Comparison to Boghossian’s View

The critical point of section III rests largely on the claim that dispositions are cognitively holistic. Paul Boghossian also parlays the point about holism into a criticism of Fodor’s theory of content³¹ and into a criticism of the appeal to dispositions in naturalistic semantics in general.³² In this section, I review Boghossian’s criticism of Fodor’s theory of content and argue that the criticism I have developed is the more potent of the two: Boghossian’s criticism requires

that we pin upon Fodor's theory a very specific and controversial implication, while my critical approach does not.

In *TOC*, Fodor bemoans the use of a common, but by his lights unproductive, strategy in naturalistic semantics.³³ The typical naturalistic semantic theory separates tokenings of a given mentalese term into two groups: those that occur in reference-determining (type one) situations and those that occur in non-reference-determining (type two) situations. Dretske's talk of a learning period well illustrates the distinction between type one and type two situations. Tokenings of given mentalese term *t* that occur during the learning period are type one tokenings and determine *t*'s extension: it consists of whatever is of the same kind as that about which the type one tokenings of *t* carried information. After the learning period ends, tokenings are of type two: *t*'s extension is fixed; that *t*'s tokening is now sometimes caused by items not in its extension in no way alters that extension.³⁴ We can see, then, how the type one/type two distinction might explain the occurrence of misrepresentation: Items not in the extension of mentalese term *t* sometimes cause type two tokenings of *t*; these tokenings count as instances of misrepresentation, because the circumstances surrounding the subject's type one tokenings of *t* (however a specific theory characterizes these cases) independently fixed *t*'s extension in such a way as to exclude the items in question.

By an argument I do not rehearse here, Boghossian concludes that Fodor himself is committed to the content-determining power of type one situations: Boghossian claims that if Fodor's theory of content is to assign the correct extension to a given mentalese term, there must be at least one possible world where only the members of the term's correct extension can cause the tokening of that term. However, given the cognitively holistic nature of dispositions, it seems that in any world you choose, a subject's current tokening-mediating beliefs might cause

her to token the term in question in response to something that should not be in the term's extension. In order to guarantee the appropriately content-fixing, type one world, Fodor must, in moving to this world, strip the subject of the variety of mental states that could cause her to token the term in question in response to something other than a member of what we take to be the term's proper extension; yet Fodor cannot accomplish this without begging the question, i.e., without assuming he has already in hand some way to characterize the content of the relevant beliefs. For example, if Fodor's theory of content is to assign horses and only horses to the extension of 'horse', there must exist a world where only horses cause the tokening of 'horse'. For all we know, though, the subject in question holds the two following beliefs: "today is Tuesday" and "whenever it's Tuesday, I should think about horses in response to anything I see." If we set the subject down in what we think is a type one world, things other than horses will cause her to token 'horse', the world in question thus failing to accomplish its reference-fixing work. For Fodor's theory to yield the correct extension assignment, Fodor must somehow specify that the subject is not in any of the mental states that might cause her to token 'horse' in response to something other than a horse. This creates an insurmountable obstacle to Fodor's naturalistic approach: he set out to completely reduce mental content to nonsemantic facts; however, it now seems that if he is to specifically exclude, as he must, the potentially problematic mental states from his specification of the content-determining conditions, he will have to characterize those states in terms of their content.³⁵

My critical approach differs significantly from Boghossian's. Whereas Boghossian's criticism implies Fodor's commitment to type one content-fixing situations, my approach does not. This is of genuine moment, given Fodor's substantial arguments that he is not committed to the existence of type one content-fixing situations.³⁶

Recall the basic claim of Fodor's theory of content: Mentalese term 'A' refers to As if all nomic connections between non-As and 'A' are asymmetrically dependent on the connection between As and 'A'. In other words, for any non-A that sometimes causes the tokening of 'A', the following holds. If As were to no longer cause the tokening of 'A', then *ceteris paribus*, the non-As would not cause the tokening of 'A', but not vice versa: if the non-As were to no longer cause the tokening of 'A', then, *ceteris paribus*, As *would* continue to cause the tokening of A. This pattern of dependencies establishes the connection between As and 'A' as the most fundamental of all such connections.

In essence, Boghossian claims that Fodor must work the trick of nomic dissociation all at once: in order to fix the reference of 'A', there must be a world where *all* of the non-As are nomically dissociated from 'A', while A's continue to cause 'A'. Fodor rejoins that he need not; instead, he can make comparisons of two nomic connections at a time: to show that the connection between horses and 'horse' is more fundamental than that holding between cows on dark nights and 'horse', Fodor claims he need only ask you to imagine what happens when horses no longer cause the tokening of 'horse' and, separately, what happens when cows on dark nights no longer do. Even if buffalo at a great distance sometimes cause subjects to token 'horse', that is a matter to be handled separately: by *independent* thought experiment we establish that the horses-cause-'horse' connection is more fundamental than the causal connection between buffalo at a distance and 'horse'.³⁷

Without attempting to adjudicate the dispute between Boghossian and Fodor, I note only that the criticism of Fodor's theory set forth in section III does not imply a decision in favor of either party: I need not charge Fodor with a commitment to type one situations, where only the members of a term's extension cause its tokening. Fodor claims that his theory assigns the right

extensions so long as we compare two nomic connections at a time and then take the overall pattern of results into account. I accept Fodor's presentation of the theory but claim that the cognitive holism of dispositions sabotages his view nevertheless. Take two nomic connections, ignoring all others. Sever one and then the other, in order to see which connection is asymmetrically dependent on the other. Still, no matter how the universe is altered to sever one of the lawlike connections, some of the subject's beliefs may, upon this alteration, inhibit display of the supposedly content-fixing pattern of tokenings: as in Oscar's case from section III, these beliefs may cause the wrong connection to appear to be, by Fodor's test, the more fundamental of the two. This renders my criticism more effective than Boghossian's: I avoid the controversy over Fodor's alleged commitment to type-one situations and criticize the theory as Fodor himself describes it.

V. An Update

In *Concepts*, Fodor concerns himself primarily with issues related to concept structure and concept acquisition. When Fodor discusses conceptual content, he does so, for the most part, in nonspecific terms: he endorses informational semantics (the view that patterns of nomic relations fix conceptual content) as a general theoretical approach. However, at certain points, Fodor significantly amends his views about content, by way of discussing concept possession.³⁸ In this section, I argue that despite these amendments, the problem caused by the cognitively holistic nature of dispositions remains.

Fodor now unequivocally accepts a purely informational theory of conceptual content. Such a theory, as Fodor understands it, founds content--a concept's 'locking to' a kind or property--on subjects' dispositions;³⁹ this alone would seem to bestow upon Fodor's current view the same shortcoming that afflicts the theory of content presented in *TOC*. Recall, though, that the force

of my criticism in section III depended largely on Fodor's prior commitment to semantic atomism as regards natural kind terms in mentalese. In *Concepts*, Fodor mitigates his atomism in a way that appears to notably alter the dialectic. Fodor now allows that the typical, perhaps even every, human employs such mentalese terms as 'hidden essence', 'natural kind', and 'microstructure' as means by which to acquire everyday natural kind concepts in the full-blown way that twentieth-century Oscar has acquired 'water.'⁴⁰ Given Fodor's claim that possession of 'hidden essence', 'natural kind', and 'microstructure' serves as a prerequisite of sorts for the possession of everyday natural kind concepts, he no longer seems committed to saying that the content of 'water' is fixed atomistically.

To some extent, dispositions persist in their mischief, for Fodor remains committed to the metaphysical truth of semantic atomism. In *Concepts*, Fodor allows that a being who does not possess any such concepts as 'hidden essence', 'microstructure', or 'natural kind' could, *in principle*, achieve full-blown reference to a natural kind such as H₂O. Successful reference requires only that the subject lock to, i.e., bear the right nomic relation to, the natural kind in question; and thus, Fodor says, a subject can, in theory, lock to a natural kind atomistically.⁴¹ For Fodor, locking to a property or kind amounts to no more than instantiating the appropriate nonsemantically individuated disposition; therefore, as a metaphysical theory of content, Fodor's current view does not overcome the objection raised above in section III.

Fodor appears to face problems even in the typical twentieth-century human's case, for whom the grounding of 'water', as a full-blown natural kind term, does not proceed atomistically. Consider how the nonatomistic view that Fodor presents in *Concepts* would seem to answer the difficulty raised in section III. To achieve reference to water, modern humans direct certain intentions toward 'water', intentions themselves consisting of concepts such as

‘hidden essence’, ‘microstructure’, and ‘natural kind’. Thus, by dint of Oscar’s intention alone, Fodor can freely attribute to Oscar a concept ‘water’ that determinately refers to H₂O.

Regardless of Oscar’s beliefs (e.g., suspicions directed toward chemists) and dispositions, the content of one of his intentional states, together with facts about his local environment, fixes the content of ‘water’. When he tokens ‘water’ he intends to refer only to stuff of the same natural kind as (i.e., with the same hidden essence as) local samples, and given his historical environment, this intention determines that his mentalese term ‘water’ refers to H₂O and not XYZ.

I shall press two complaints against this nonatomistic view of content fixation. First, notice that the preceding story exonerates Fodor’s theory of content by abandoning it. Asymmetric dependence and dispositions to token in counterfactual circumstances now play no role in fixing the reference of ‘water’; the intention to refer only to stuff of the same kind as local samples does all of the work. Perhaps militia Oscar’s intention that ‘water’ refer only to stuff of the same kind as local samples is enough to fix the content of ‘water’ for him, but it does nothing to establish the pattern of counterfactual responses that Fodor’s asymmetric dependence-based theory of content requires. Oscar may intend that ‘water’ refer only to stuff of the same kind as local samples; however, because of the cognitively holistic nature of his dispositions, he still may not token ‘water’ in response to H₂O in worlds where he can discriminate between H₂O and XYZ. Simply return to the story about the chemists and their coding scheme.

Fodor will likely respond to my criticism by charging that I have misrepresented the way in which possession of ‘microstructure’ and the like contributes to the fixation of content for typical natural kind terms. Fodor’s own story differs significantly from the one I have just told. The modern human does not fix the content of ‘water’, for example, simply by intending to refer to

the natural kind whose members share their microstructural properties with local samples; the process is much more complicated. As Fodor sees it, only as humans develop science do ‘microstructure’ and the like make their distinctive contribution to the fixation of reference for everyday natural kind concepts, such as ‘water’. It is only then that humans’ typical natural kind concepts come to refer to natural kinds robustly, across all possible worlds. Humans gain this accomplishment by explicitly deploying ‘microstructure’ and its ilk in such a way as to establish the appropriate content-fixing nomic relations: When humans notice that samples appearing to be of two different types have the same effect on samples of a third type, humans begin to suspect the influence of hidden essences or microstructures; humans then contrive experiments (i.e., they do some science) intended to get themselves into the right nomic relations to the underlying essences or microstructures of the kinds in question, which kinds humans had previously thought about only in terms of their superficial properties.⁴²

Fodor’s more complicated story does little to get him off the hook, primarily for reasons already explained: Even after humans develop science, they do not stand in the nomic relations to everyday natural kinds that Fodor’s theory of content requires if correct assignments of reference are to be made. The argument of section III shows that neither Oscar nor the typical twentieth-century scientist instantiates the “appropriate” dispositions to token ‘water’, the dispositions that would according to Fodor’s theory of content, determine H₂O to be the extension of the mentalese term ‘water’.⁴³ The present complaint takes the form of a dilemma for Fodor. Either (1) he opts for a nonatomistic story as I’ve told it, where it is the content of the intention to treat a natural kind term as such that fixes reference, or (2) he cleaves to *Concepts*’ nonatomistic story of reference-fixation for everyday natural kind terms. In the former case, he

compromises his asymmetric dependence theory of content, for patterns of counterfactual tokenings do not fix reference; in the latter case, his theory assigns the wrong extensions.

I introduce my second complaint by asking a question of Fodor: How do humans come to possess the very concepts ‘microstructure’, ‘hidden essence’, and ‘natural kind’? Here Fodor faces another dilemma. His theory of content-fixation for ‘microstructure’ and the like cannot be the same as the theory he presents in *Concepts* to explain how content is fixed for humans’ typical natural kind concepts; the latter theory assumes that humans already possess ‘microstructure’ *et alia*, employing such concepts as means by which to ground everyday natural kind concepts. Of course, the concept ‘microstructure’ itself cannot be used as means by which to acquire the concept ‘microstructure’. Thus, either Fodor must treat ‘microstructure’ and the like as having their content fixed atomistically, by subjects’ nonsemantically individuated dispositions alone, in the way he treats typical natural kind terms in *TOC*; or he must provide another theory of content appropriate to ‘microstructure’ and its ilk. In what follows, I treat the cases in order.

Assume Fodor chooses the first route, accepting that subjects’ dispositions atomistically fix the reference of ‘microstructure’. If Fodor proceeds in this way, he subjects his view to the same problem that arose out of the earlier discussion of Twin Earth cases. Scientifically sophisticated thinkers are presumed to have certain dispositions in respect to ‘microstructure’, in virtue of which ‘microstructure’ tracks the kind *microstructure*; however, such dispositions are no less cognitively holistic than Oscar’s. One might wonder whether the present case is analogous to Oscar’s, whether ‘microstructure’ can misrepresent in the way that ‘water’ can; but surely this is a possibility: Consider such tokenings of ‘moved by internal microstructural effects’ as a magician’s trick might cause, where, say, what appears to be a self-locomoting structure turns

out to be an empty shell, cleverly manipulated from outside. On the present assumption, Fodor must explain such misrepresentation by appeal to dispositions to differentially token ‘microstructure’ in cases where scientifically sophisticated subjects can discriminate between real microstructures at work and effects for which objects’ microstructures are not directly responsible. As it was in Oscar’s case, dispositions cannot do the work Fodor requires of them. Imagine a nearby possible world in which moving empty shells no longer cause the tokening of ‘microstructure’ in a given scientist; perhaps magicians have revealed their tricks to all the world. Fodor faces the same problem as arose in militia Oscar’s case: whether the scientist will exhibit the “correct” pattern of tokenings (i.e., tokening ‘microstructure’ in response to microstructures at work but not in response to, for instance, empty shells being moved by strings) depends on her other beliefs; the scientist might have a standing belief that she should never accept anything a magician says. If the world is to become such that a scientifically sophisticated subject can now recognize a heretofore undetectable difference between the effects and noneffects of microstructure, the universe must be altered; and in respect to whatever alteration we imagine, this subject might, for all we know, have some belief that causes her pattern of tokenings to be something other than what Fodor’s theory of content tells us to expect from someone whose mentalese term ‘microstructure’ refers only to microstructures.

Moving on to the second case: Here Fodor must offer something other than dispositions, nonsemantically individuated, to fix the content of ‘microstructure’ and the like. I locate two possibilities in Fodor’s work, and neither seems likely to explain in a satisfactory way how the content of such terms as ‘microstructure’ is fixed. First consider Fodor’s treatment of the logical vocabulary: Fodor claims that functional roles alone fix the content of these terms.⁴⁴ We should not dismiss out of hand the possibility that functional roles alone define ‘microstructure’, ‘hidden

essence', and 'natural kind'; perhaps the content of 'natural kind', for example, is fixed entirely by the patterns of inference developmental psychologists look for as evidence that children possess the concept of a natural kind (one such pattern is exemplified by the child's refusing to believe that doctoring a raccoon to make it appear to be a skunk turns the raccoon into a real skunk).⁴⁵ Note, however, that 'microstructure' and the like differ from the logical vocabulary in a way that should make naturalists, Fodor especially, hesitant to accept a functionalist theory of content for the former terms. Functional-role semantics, says Fodor, leads to holism (and naturalists should shudder at the thought); inferential connections constitute at least part of a term's functional role, and for most mentalese terms, there is no principled way to limit the number of inferential connections that constitute the term's definition: trace inferential connections far enough, and you see that every term is connected to every other, or so it would seem. However, the situation differs for the logical vocabulary: it seems plausible that a fairly limited number of inferential connections exhaustively defines terms in the logical vocabulary: being the term 'and' just amounts to such things as being the term that appears between two propositions each of which the subject is inclined to accept when it stands alone; in contrast it is not at all clear that 'natural kind', 'microstructure', and 'hidden essence' are *exhaustively* definable by a distinctive and manageable set of inferential connections. At least Fodor offers no argument for such a position.

An alternative remains.⁴⁶ In *Concepts*, Fodor identifies concepts of mind-dependent properties as a type of concept that differs both from true natural kind concepts and mentalese terms in the logical or mathematical vocabulary.⁴⁷ It seems possible, then, that Fodor intends 'microstructure' and the rest to be concepts of mind-dependent properties and that they are to have their content fixed in a way peculiar to such concepts. Alas, we find no distinctive theory

of content for ‘microstructure’ and its ilk in Fodor’s discussion of mind-dependent properties,⁴⁸ for Fodor’s theory of content for concepts that refer to mind-dependent properties is informational. According to Fodor, a concept has a mind-dependent kind or property as its content when the relevant subject’s concept locks to a mind-dependent kind or property, and locking to is constituted by the subject’s possessing the right dispositions; thus, Fodor’s theory of content for concepts of mind-dependent kinds and properties would, even were Fodor to apply it to ‘microstructure’ and the like, be subject to all of the problems attendant upon Fodor’s other appeals to nonsemantically individuated dispositions.

I have critically reviewed Fodor’s theory of content for natural kind terms in mentalese, concluding that Fodor’s theory faces a crippling problem: Subjects’ dispositions to token mentalese terms in counterfactual situations are supposed to fix reference, yet those dispositions are cognitively holistic. Thus, Fodor’s theory cannot assign the extensions we should expect, unless he abandons his atomism. And although Fodor mitigates his atomism in his more recent work, such mitigation only solves the problem, if it does at all, by abandoning his theory of content, deferring instead to the contents of subjects’ intentions.

From my critical appraisal of Fodor’s views, I draw a broader moral. In explaining the fundamental basis of content, a naturalistic semantic theory for mentalese terms should not attempt to reduce reference-fixation to subjects’ dispositions to token mentalese terms under counterfactual circumstances; it is difficult to see how any such theory will avoid the disruptive effects of the cognitively holistic nature of dispositions. This broader conclusion does not spell disaster for naturalistic semantics; the naturalist may choose to pursue one of the other available naturalistic options. I recommend that naturalists take seriously the sort of theory with which

Fodor flirts in *TOC* and eventually rejects: a theory that founds reference, at the fundamental level, on the subject's actual interactions with her environment. I have fleshed out this approach elsewhere as part of a hierarchical theory of content-fixation: at the basic level, the subject's actual history determines the reference of her natural kind terms in mentalese (typically this kind of content-fixation occurs early in human development); content so-fixed helps to fix the reference of further mentalese terms, via various of the subject's reference-fixing intentions.⁴⁹ Alternatively, many naturalists propose attending to the evolutionary history of the *species* as the ultimate ground of reference.⁵⁰ However, the tasks of developing the range of remaining naturalistic options and adjudicating among them lie beyond the scope of the present work.

Humanities Division

Green River, Tacoma, and Highline Colleges

¹ I benefited from discussing some of the issues addressed in the present work with Charles Chastain and Walter Edelberg.

² In order to make progress in the present context, I bracket substantive questions about the nature of natural kinds. As a working notion, consider a natural kind to be any kind that plays an explanatory role in the natural sciences.

³ I move freely between the use of 'mental representations', 'concepts', and 'terms in mentalese', although I favor the last of the three. Also, I use the terms 'reference' and 'content' interchangeably; this is Fodor's practice, and given that I do not here intend to challenge the philosophical views on which this practice rests, I follow Fodor. For some of Fodor's reasons for thinking that the content of a mentalese term amounts to no more than its reference, see Fodor's *A Theory of Content and Other Essays*, chapter 6. (Cambridge: MIT, 1990) (*TOC*, hereafter).

⁴ *Psychosemantics*, p. 121. (Cambridge: MIT Press, 1987) (*PS*, hereafter).

⁵ *Concepts: Where Cognitive Science Went Wrong*, pp. 78-79. (Oxford: Oxford University Press, 1998) (*Concepts*, hereafter).

⁶ *TOC*, p. 92. Fodor qualifies his atomism in *The Elm and The Expert*, pp. 75-76 (Cambridge: MIT Press, 1994) (*EATE*, hereafter) by allowing that determinate reference requires the possession of logical terms such as 'and' and

'not'; this he does in order to solve Quine's problem of the indeterminacy of reference. However, the presupposition of logical vocabulary does not, in itself, lessen Fodor's commitment to the atomic status of natural kind terms *with respect to each other* in the fixation of their content; it is this commitment that is of present concern. In *Concepts* (p. 156), Fodor qualifies his atomism in a way that would seem to bear more directly on the present discussion, more on which below.

⁷ For Fodor's talk about dispositions, see *EATE*, pp. 31-32; *Concepts*, pp. 77, 155; for what appear to be unqualified endorsements of counterfactual talk, see *EATE*, p. 115; *Concepts*, p. 157.

⁸ *TOC*, pp. 95, 101; *Concepts*, p. 145, n. 18.

⁹ I have little to say here about the proper understanding of naturalism. For present purposes, naturalism need only imply that we can, in principle, replace all accurate descriptions of the world stated in intentional idioms with accurate descriptions couched in nonintentional terms (though even in respect to this point, we find dissenting naturalists; see, for example, L. R. Baker, *Explaining Attitudes* (Cambridge: Cambridge University Press, 1995), and L. Shapiro, "Representation from Bottom and Top", *Canadian Journal of Philosophy*, XXVI, no. 4 (Dec. 1996): 523-42.)

¹⁰ Talk of asymmetric dependence is Fodor's way of explaining the priority of certain causal connections in fixing reference. A causal connection between a natural kind and a term in mentalese is asymmetrically dependent on the causal connection between a second natural kind and that same mentalese term if and only if breaking the first connection leaves the second intact, but not vice versa. For Fodor's presentation of the asymmetric dependence theory, see *PS*, Chap. 4, and *TOC*, Chap. 4. For critical discussions, see H. Cram, "Fodor's Causal Theory of Representation," *Philosophical Quarterly*, 42 (1992): 56-70; M. Gibson, "Asymmetric Dependencies, Ideal Conditions, and Meaning," *Philosophical Psychology*, 9 (1996): 235-259; P. A. Manfredi and D. M. Summerfield, "Robustness without Asymmetry: A Flaw in Fodor's Theory of Content," *Philosophical Studies*, 66: 261-283; W. Seager, "Fodor's Theory of Content: Problems and Objections," *Philosophy of Science*, 60: 262-277; and various chapters of B. Loewer and G. Rey (eds.), *Meaning in Mind: Fodor and His Critics* (Oxford: Blackwell, 1991).

In what follows, I evaluate Fodor's theory of content as a theory of content for humans' mentalese terms. This seems to run counter to Fodor's own characterization of his project: sometimes Fodor claims only to be giving sufficient conditions for intentionality, in the attempt to show that Brentano is wrong to claim that intentionality is irreducible (*TOC*, p. 96). If the asymmetric dependence theory is only supposed to state a sufficient condition for

content-fixation, then its failing to provide the right answer in the human case is no strike against the theory.

However, from the naturalist's standpoint, we can only evaluate Fodor's alleged solution to Brentano's problem by applying the theory to the sole type of full-blown intentional/representational system we know of, i.e., the human being. If Fodor's theory does not yield correct answers in the case of humans, what could justify a naturalist's claim that asymmetric dependence is sufficient to fix intentional content? *A priori* insight? Setting all of these issues aside, note that it is of intrinsic interest whether Fodor's asymmetric dependence theory accurately describes the conditions for content-fixation for human concepts.

¹¹ *TOC*, p. 114.

¹² For Putnam's description of the Twin Earth cases, see "The Meaning of 'Meaning'," reprinted as chapter 12 of Putnam's *Mind, Language, and Reality: Philosophical Papers, Volume 2* (Cambridge: Cambridge University Press, 1975).

¹³ *TOC*, p. 115.

¹⁴ *TOC*, p. 115.

¹⁵ I have simplified matters a bit here. When Fodor is being careful, he limits the claim in the text to the nearest possible worlds in which H₂O and XYZ differentially affect Oscar; one world is nearer to our world than a second is to our world if and only if there are fewer differences between our world and the first than there are between our world and the second.

¹⁶ *TOC*, p. 134, n. 25.

¹⁷ *TOC*, p. 95.

¹⁸ *EATE*, p. 115.

¹⁹ *EATE*, p. 118; also see the second set of passages cited in note 6 above.

²⁰ *TOC*, p. 116.

²¹ *TOC*, pp. 55-56.

²² *TOC*, pp. 56, 99, 120.

²³ *Concepts*, p. 145, n. 18. For present purposes 'locking' can be thought of as the reference-fixing relation; more on this below, in section V.

²⁴ *Concepts*, p. 14.

²⁵ For Fodor's discussion and comparison of the two versions of his theory, see *TOC*, pp. 120-124.

²⁶ The Language of Thought, p. 101. (Cambridge: Harvard University Press, 1975).

²⁷ It is possible that the subject possesses an explicit, though not necessarily conscious, intention to treat ‘water’ as a natural kind term, and that the content of this single intention fixes the content of ‘water’. In this case, talk of amending Fodor’s theory seems quixotic: patterns of counterfactual tokenings do no reference-fixing work; it is the content of the single intention to treat ‘water’ as a natural kind concept, together with facts about the local environment, that determines ‘water’ to refer to H₂O. I consider this possibility further in the closing section.

²⁸ Confusion regarding the best way to state Fodor’s position arises out of his discussion of the Twin Earth cases. In *TOC*, Fodor seems to want to respect Putnam’s (and Kripke’s) idea that there is something special about the semantics of natural kind terms, i.e., that we fix their reference by intending to treat these terms in a certain way. However, it’s not clear why Fodor needs to tell any special story about Twin Earth and natural kind terms: one might expect Fodor to say only that his theory of content solves the Twin Earth puzzles because the asymmetric dependencies work out the way that they’re supposed to. This might not seem to fairly reflect Fodor’s view, given his explicit contrast between the way contents are fixed for terms in the logical and mathematical vocabulary and the way contents are fixed for other terms. However, Fodor does not intend his theory of content to apply to the former terms (*TOC*, pp. 110-111), and thus I can safely ignore his remarks about them. The present question is whether, *within the purview of Fodor’s theory of content*, we are to give natural kind terms any special treatment.

²⁹ In his discussion of the variety of stimuli that might cause a subject to token the mentalese term ‘cow’, Fodor himself notes, “Since, after all, cow spotting can be mediated by theory to any extent you like, the barest whiff or glimpse of cow can do the job for an observer who is suitably attuned.” (*TOC*, p. 109) Fodor makes similar remarks in response to Louise Antony and Joseph Levine’s criticisms of his theory of content. See Fodor’s “Replies” in Loewer and Rey, *op. cit.*, pp. 255-319, especially p. 256. Also see *Concepts*, p. 78.

³⁰ It might be argued that I have ignored a crucial possibility: perhaps we can imagine changes in the universe that render H₂O and XYZ distinguishable to Oscar, where, unlike the case of the chemists’ color-coding scheme, these changes are not the subject of any of Oscar’s beliefs. Acknowledging this possibility does not weaken the argument given in the text. First off, note that for Fodor to employ only such changes as were just described as the basis of his content assignments, he must have some way of distinguishing between the cases where Oscar has and has no beliefs concerning a given nomic change; this is problematic for reasons discussed below in section IV. Secondly, I am suspicious of the idea of nomic changes that render H₂O and XYZ distinguishable to Oscar but in respect to

which Oscar's beliefs have no possible relevance. Such changes are easiest to imagine when we think of altering the universe in ways that are not consciously detectable to Oscar, e.g., because we have changed subtle ways in which his perceptual system works. However, it seems that any change one makes here could be the subject of one of Oscar's *subconscious* beliefs or other subconscious states with representational content. And so long as Oscar possesses a subconscious, content-laden state that has some, even indirect, bearing on the nomic change in question, Fodor's theory continues to face the problem set out in the text.

³¹ "Naturalizing Content," in Loewer and Rey, *op. cit.*, pp. 65-86, especially pp. 78-83.

³² "The Rule-Following Considerations," *Mind*, 98 (1989): 507-549; especially pp. 539-540.

³³ *TOC*, p. 60.

³⁴ Knowledge and the Flow of Information (Cambridge: MIT Press, 1981).

³⁵ On the basis of the point about holism, Boghossian develops two distinct criticisms of Fodor's theory: the one on which I have focused in the text and one implied by the infinite character of the set of beliefs that might influence the tokening of a given mentalese term; in fact, when Boghossian criticizes Fodor's theory of content in particular, as opposed to naturalistic theories of content in general, he focuses on the latter; see Alexander Miller's "Boghossian on Reductive Dispositionalism about Content: The Case Strengthened," *Mind and Language* 12 (1997): 1-10. As the latter criticism has it, Fodor commits himself to doing the impossible: it is simply impossible to specify the type one situations Fodor's theory requires, because the specification must exclude from the subject's mind the infinite number of beliefs that could mediate the tokening of a given mentalese term in her. My criticism of Fodor's theory of content is more similar to the first of Boghossian's criticisms than it is to the second of his criticisms; thus I discuss only the first criticism in detail. However, the reason I give in the text for preferring my tack over Boghossian's also applies when comparing my approach to Boghossian's second criticism: in each case, the cogency of Boghossian's criticism depends on his first showing that Fodor's theory is committed to type one situations.

³⁶ See Fodor's reply to Boghossian in Loewer and Rey, *op. cit.*, pp. 271-277.

³⁷ In his critical discussion of holism, Fodor takes seriously (i.e., does not view as outright crazy) the possibility of a punctate mind: a mind that possesses only one concept; see J. Fodor and E. Lepore, Holism: A Shopper's Guide (Oxford: Blackwell, 1992). Furthermore, Fodor supposes that we can, in principle, functionally characterize belief as a mental state type (*PS*, p. 17). Combining these ideas, Fodor might pursue a different response to Boghossian:

admitting the need for type one situations as Boghossian describes them, Fodor could claim that the content-determining counterfactual situations are those in which the subject has been stripped of all beliefs, possessing only the two concepts whose relative dependence we are out to evaluate (here we conjecture not quite a punctate mind, but one that only a committed atomist such as Fodor will feel comfortable imagining). If Fodor were to take this approach, he would not beg the semantic question, for he would need only describe belief *as a state type*, without mentioning the *contents* of any beliefs; Fodor could describe the requisite type one situation by saying something of the following form: there does not exist in the subject an x such that the nonsemantic functional characterization of belief as a general state-type is true of x . However, if, in order to save his theory of content, Fodor must make this move, he does so at great cost. For Fodor's theory to be of any interest to naturalists, it should apply to at least some actually existing representational systems; but as regards the representational systems with which we are familiar, we have no idea which mentalese terms these systems will token in counterfactual circumstances where they have been stripped of all of their beliefs.

³⁸ Sometimes Fodor claims that to possess a concept p is just to have mentalese term with the content p (*Concepts*, p. 124), in which case, whatever Fodor says about conditions for possessing concept p applies equally to the conditions for a concept's having the content p . When Fodor is more careful, he leaves open the possibility that there is something more to concept possession than having a mentalese term with the right content: “[H]aving a concept (concept possession) is constituted, *at least in part*, by *being in* some sort of nomic, mind-world relation.” (*Concepts*, p. 121, first italics added) and “[C]ontent individuation can't be all there is to *concept* individuation.” (*Concepts*, p. 15) The concept ‘H₂O’ differs from the concept ‘water’, the former, but not the latter, being syntactically complex; this is the case even though, for the typical twentieth-century subject, the two concepts have the same content. For some purposes, it is essential that Fodor distinguish between concept individuation and issues related only to conceptual content: sometimes we best explain a subject's behavior by taking into account the different inferential roles of concepts that share the same content (see *TOC*, chapter 6). For present purposes, however, we need not worry much about the added difference that distinguishes concepts having the same content. These extra conditions of concept individuation do not bear on conceptual content; we are presently interested in Fodor's theory of content, not his views about psychological explanation. Thus, in the later chapters of *Concepts*, when Fodor talks about the nomic relations that underlie concept possession, we can reasonably take whatever he

says to have direct implications for his theory of content--especially in so far as what he says about concept possession might render the theory of content more defensible.

³⁹ *Concepts*, p. 126.

⁴⁰ *Concepts*, p. 156.

⁴¹ *Concepts*, p. 158.

⁴² For Fodor's story of how humans work their way up to the acquisition of full-blown natural kind concepts from more superficial representations of the world, see *Concepts*, 150-161.

⁴³ Fodor's story seems implausible for a second and important reason. In contrast to the view of many developmental psychologists, Fodor does not believe that young humans naturally do what it takes to refer to natural kinds as such. Here Fodor dismisses far too quickly developmental research suggesting that young children think in terms of hidden essences and thus acquire full-fledged natural kind concepts without having to do science as it's normally conceived of (*Concepts*, pp. 154-155). While Fodor is willing to allow that recent developmental work establishes young children's sensitivity to underlying structure (or to an appearance/reality distinction), Fodor demands evidence that children make the proper *causal* connection, intending to refer to whatever underlying structure is causally responsible for a kind's superficial properties. Fodor's demand is fair enough; however, there is some reason to think his demand can be met. For evidence of causal reasoning in very young children, see P. J. Bauer and J. M. Mandler, "One Thing Follows Another: Effects of Temporal Structure on 1- to 2-Year-Olds' Recall of Events", *Developmental Psychology*, 25 (1989): 197-206; for evidence of causal reasoning in three- and four-year-olds, see C. M. Massey and R. Gelman, "Preschooler's Ability to Decide Whether a Photographed Unfamiliar Object Can Move Itself," *Developmental Psychology*, 24 (1988): 307-317; R. Gelman, "First Principles Organize Attention to and Learning about Relevant Data: Number and the Animate-Inanimate Distinction as Examples," *Cognitive Science* 14 (1990): 79-106; for an extended argument that young children develop theories of the world in much the same fashion as mature scientists, see A. Gopnick and A. N. Meltzoff, Words, Thoughts, and Theories. (Cambridge: MIT Press, 1997).

⁴⁴ *TOC*, pp. 110-111.

⁴⁵ For details of such transformation experiments, see F. Keil, Concepts, Kinds, and Cognitive Development, chapters 9-11. (Cambridge: MIT Press, 1989).

⁴⁶ And as Fodor sees things now, consideration of this alternative exhausts all possibilities; see *Concepts*, p. 141.

⁴⁷ *Concepts*, chapters 6 and 7, *passim*.

⁴⁸ Furthermore, we should wonder how it is possible that content could ever be fixed for concepts of mind-dependent properties. A mind-dependent property is one whose objective status depends on the fact that humans lock to the property under certain circumstances (*Concepts*, p. 137); here Fodor's views seem caught in circularity. Informational semantics assigns content on the basis of nomic relations, which presumably have to hold between real properties. However, a mind-dependent property is not a real property except that people lock to it. But if it's not a real, objective property, how, you might wonder, does anyone ever lock to it in the first place? The property would seem to require independent status in order for the right causal relations to come to hold between a subject and the property. The property cannot both achieve its ontological status because we lock to it and exist *before* we lock to it.

⁴⁹ R. D. Rupert, "The Best Test Theory of Extension: First Principle(s)," *Mind and Language* 14 (1999): 321-355; "On the Relationship between Naturalistic Semantics and Individuation Criteria for Terms in a Language of Thought," *Synthese* 117, 1 (1999): 95-131.

⁵⁰ See R. G. Millikan, Language, Thought, and Other Biological Categories (Cambridge: MIT Press, 1984) and the great majority of work done under the rubrics 'teleosemantics' and 'teleological theories of content'. For a review of some of this work, see K. Sterelny, The Representational Theory of Mind: An Introduction, chapter 6. (Oxford: Blackwell, 1990). For more recent work, see C. L. Elder, "What versus How in Naturally Selected Representations," *Mind*, 107 (1998): 349-363; K. Neander, "Misrepresenting and Malfunctioning," *Philosophical Studies*, 79 (1995): 109-141; D. Papineau, "Teleosemantics and Indeterminacy," *Australasian Journal of Philosophy*, 76 (1998): 1-14; M. Rowlands, "Teleological Semantics," *Mind* 106 (1997): 279-303; L. A. Shapiro, "Representation from Bottom and Top," *Canadian Journal of Philosophy* 26 (1996): 523-542; and many of the works cited in those pieces. For critiques of various versions of the teleological approach, see R. D. Rupert, "Mental Representations and Millikan's Theory of Intentional Content: Does Biology Chase Causality," *Southern Journal of Philosophy* 37 (1999): 113-140; "The Best Test Theory of Extension: First Principle(s)," *Mind and Language* 14 (1999): 321-355, section 2.2; *TOC*, chapter 3; and R. Cummins, Representations, Targets, and Attitudes, pp. 45-47. (Cambridge: MIT Press, 1996).